Children across societies enforce conventional norms but in culturally variable ways

Patricia Kanngiesser[a,b,1], Marie Schäfer[c], Esther Herrmann[d], Henriette Zeidler[e], Daniel Haun[f], and Michael Tomasello[g]

[a]School of Psychology, University of Plymouth, Plymouth PL4 8AA, United Kingdom; [b]Faculty of Education and Psychology, Freie Universität Berlin, 14195 Berlin, Germany; [c]Department of Philosophy, Universität Leipzig, Leipzig, Germany; [d]Department of Psychology, University of Portsmouth, Portsmouth PO1 2DY, United Kingdom; [e]School of Psychology, Aston University, Birmingham B4 7ET, United Kingdom; [f]Department of Comparative Cultural Psychology, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany; and [g]Department of Psychology and Neuroscience, Duke University, Durham, NC 27708

Edited by Patricia Kuhl, Institute for Learning & Brain Sciences, University of Washington, Seattle, WA; received July 7, 2021; accepted November 18, 2021

Individuals in all societies conform to their cultural group's conventional norms, from how to dress on certain occasions to how to play certain games. It is an open question, however, whether individuals in all societies actively enforce the group's conventional norms when others break them. We investigated third-party enforcement of conventional norms in 5- to 8-y-old children (*n* = 376) from eight diverse small-scale and large-scale societies. Children learned the rules for playing a new sorting game and then, observed a peer who was apparently breaking them. Across societies, observer children intervened frequently to correct their misguided peer (i.e., more frequently than when the peer was following the rules). However, both the magnitude and the style of interventions varied across societies. Detailed analyses of children's interactions revealed societal differences in children's verbal protest styles as well as in their use of actions, gestures, and nonverbal expressions to intervene. Observers' interventions predicted whether their peer adopted the observer's sorting rule. Enforcement of conventional norms appears to be an early emerging human universal that comes to be expressed in culturally variable ways.

norms | conventions | sanctions | coordination | cross-cultural

Norms regulate how members of a group ought to behave and enable social cohesion, coordination, and large-scale cooperation (1–3). Compliance with norms depends, among other things, on formal and informal mechanisms for sanctioning those who deviate from established norms (4–6). To date, the study of how group members in different societies sanction each other in informal ways has primarily focused on third-party enforcement of social and moral norms,* including fairness norms about resource distribution (9–14), cooperative norms in cattle raids (15, 16), ownership norms (17), and norms against harming others (17, 18). Violations of social and moral norms impact others in significant ways and so, often evoke strong emotional reactions and enforcement against violators (19).

Social and moral norm violations, however, do not represent the full range of norm violations that people encounter in their everyday lives (20). Societies also have a plethora of conventional norms: rules that determine how people ought to dress, greet each other, eat, play certain games, and behave in certain public fora like funerals or weddings (21). In contrast to moral and social norms, conventional norms coordinate behavior in arbitrary ways and within some constraints (22), can be altered by consensus or authorities (8, 23). For example, shaking hands, bumping fists, or bowing to each other all achieve the same goal of coordinating how we greet each other. The reason why we follow one rule or the other is because this is how "we" do things as a group. As conventional norms involve conformity mostly for its own sake, they present an ideal test case for the extent to which different communities informally sanction or tolerate rule transgressions.

Some studies have found that young children actively enforce conventional norms by protesting, correcting others, or reminding them of the rules, even as uninvolved third parties (24–32). However, this research has been carried out exclusively with children from middle-class families in Europe and North America, and there exists, to date, no systematic cross-cultural study on conventional norm enforcement and its intercultural variability. There is some indication that children's reasoning about conventional rules differs between societies (33, 34). Moreover, studies with adults have found cross-national variation in ratings about norm compliance and tolerance of deviant behaviors (i.e., so-called "loose" and "tight" cultures) (35) as well as in judgements about how to respond appropriately to conventional norm violations (i.e., metanorms about sanctions) (20). These previous studies, however, used ratings or judgements of hypothetical scenarios and provide only limited insight into actual behavior—children and adults may think that conventional norms should be enforced in a certain way but may not take the risk of actually sanctioning others. In addition, these studies have mostly taken place in urban, large-scale societies, and there is evidence from cross-cultural research on costly third-party enforcement of fairness norms that use of informal sanctions increases with

Significance

Humans, as compared with other animals, create and follow conventional norms that determine how we greet each other, dress, or play certain games. Conventional norms are universal in all human societies, but it is an open question whether individuals in all societies also actively enforce conventional norms when others in their group break them. We show that 5- to 8-y-old children from eight highly diverse societies enforced conventional norms (i.e., game rules) when they observed a peer who apparently broke them. Magnitude and style of enforcement varied across societies. Third-party enforcement of conventional norms appears to be a human universal that is expressed in culturally variable ways.

Author contributions: P.K., M.S., E.H., H.Z., D.H., and M.T. designed research; P.K., M.S., and E.H. performed research; P.K. analyzed data; M.S., E.H., H.Z., D.H., and M.T. provided comments and critical feedback; and P.K. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: patricia.kanngiesser@plymouth.ac.uk.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2112521118/-/DCSupplemental.

Published December 28, 2021.

*Research fields and traditions differ in their definitions of moral and social norms. For example, some classify prosocial sharing norms as social norms (4, 7) and others as moral norms (8). We thus broadly group these types of norms together as social and moral norms.

Table 1. Overview of participants and societies in the study

Population	Country	Environment	Economic base	Same rule dyads (female)	Rule conflict dyads (female)	Average age (y)
ǂAkhoe Hai‖om (ethnic group)	Namibia	Rural	Gathering, some wage work	18 (10)	18 (10)	6.6
Kikuyu (ethnic group)	Kenya	Rural	Agriculture, wage work	24 (10)	24 (10)	6.6
La Plata (urban location)	Argentina	Urban	Wage work	24 (12)	24 (12)	6.5
Leipzig (urban location)	Germany	Urban	Wage work	24 (12)	24 (12)	6.5
Pune (urban location)	India	Urban	Wage work	24 (12)	24 (12)	6.5
Quechua (ethnic group)	Bolivia	Rural	Agriculture, wage work	26 (14)	26 (14)	6.5
Samburu (ethnic group)	Kenya	Rural	Pastoralism, wage work	24 (12)	24 (12)	6.5
Wichí (ethnic group)	Argentina	Rural	Craft production, some wage work	24 (12)	24 (12)	6.5
All populations				188 (94)	188 (94)	6.5

community size (12, 14), likely because small-scale communities often rely on other mechanisms such as direct reciprocity and reputation management (36). It is an open question whether small-scale communities also engage in less third-party enforcement of conventional norms than large, urban populations, and it is possible that egalitarian hunter-gatherer communities may be particularly tolerant of conventional transgressions (37) and not show any enforcement.

We, therefore, conducted a comprehensive study with 376 (5- to 8-y-old) children from eight diverse small- and large-scale societies (Table 1 and *SI Appendix*). The study included children from three urban locations on three different continents (South America, Europe, Asia) and from five rural locations on two continents (South America, Africa). The sites differed substantially in community sizes, ranging from rural dwellings of a few hundred people to cities with millions of inhabitants; in the languages spoken; and in their economic activities, including wage labor, agriculture, and (recently sedentized) hunter-gatherers. This resulted in a maximally diverse sample (13, 38–40) and ensured diversity in both rural and urban locations, avoiding simplistic dichotomies of urban Global North vs. rural Global South (41). Children were slightly older than children in previous norm enforcement studies in Europe (28, 30) but fell within the age range in which societal differences in normative behaviors such as sharing behavior and respect for ownership emerge (38–40, 42, 43).

We introduced children to a new sorting game with two (incompatible) conventional norms (26); half of the children learned to sort by shape, and half of them learned to sort by color. Game rules are quintessential conventional norms as they are arbitrary rules that coordinate how we do things (e.g., sort objects) in a specific context[†] and because they are relatively easy to understand, have been frequently used in studies on children's norm enforcement (26, 28, 29, 44). Moreover, using a novel sorting game allowed for a well-controlled comparison across sites, as all children were equally unfamiliar with the game and were introduced to the rules in the same way. Specifically, children learned the sorting rules via a video recording of a (male) native speaker who used generic language (e.g., "This is how you play this game.") and then practiced sorting with peer partners. This wording and procedure ensured that children viewed the novel rule as a norm and not simply as an individual sorting preference (45, 46). After children had learned their respective sorting rule, one child acted as the observer, and another child acted as the player (Fig. 1). First, observers were paired with a player who had learned the same sorting rule to establish a baseline of observers' behavior in the game (same-rule condition). Then, observers were paired with a new player who had learned a different sorting rule to investigate observers' reactions to their peer's apparent norm violation (rule-conflict condition; within-subjects design with fixed order).

If children in a particular society enforce conventional norms as third parties, then we would expect higher intervention rates in the rule-conflict condition as compared with the same-rule condition (26). We expected to find societal variation in norm enforcement and speculated that children from urban, large-scale societies would show larger intervention effects than children from small-scale societies (12, 14), with the recently sedentized hunter-gatherer group (Hai‖om) showing the smallest effect (and possibly, none at all). In addition to studying whether interventions occurred, we also investigated how children enforced norms by coding children's verbal and nonverbal behaviors in detail. Norm enforcement styles or norms about how to enforce norms (so-called "metanorms") are an understudied area (20, 47), and no systematic cross-cultural comparison of actual norm enforcement behavior has been conducted to date. We expected that children in urban samples would use rule-related protest (e.g., "This is how you should do it.") more frequently than imperative protest (e.g., "Don't do it.") (28, 29, 32) but were agnostic about verbal protest preferences in small-scale societies. We were also agnostic about preferences for nonverbal intervention styles across all societies as our coding scheme of nonverbal

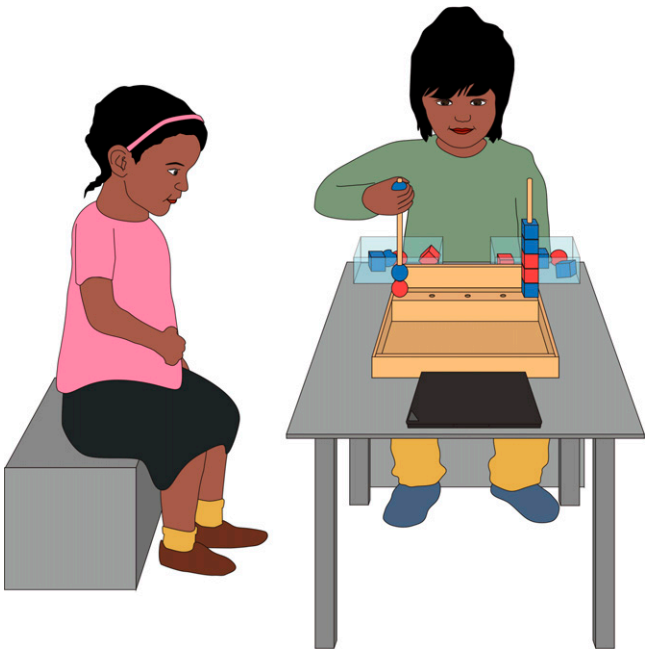

Fig. 1. Study setup. One child acted as the observer, and the other child acted as the player. Children initially learned the sorting rules via videos, featuring native speakers, from a laptop computer and then, practiced with peer partners.

[†]A binary sorting rule could be compared, for example, with the prototypical convention of what side of the road to drive on.

behaviors was more detailed than that of any previous norm enforcement study. Finally, to investigate the effectiveness of potential interventions, we analyzed whether observers' interventions in the rule-conflict condition predicted changes in players' behavior (i.e., switching to the observer's sorting rule).

Results

We coded children's nonverbal behaviors and transcribed (and translated, where necessary) their utterances from video. Coding of nonverbal behaviors included actions (e.g., putting blocks, removing them), gestures (e.g., pointing, miming), expressions (e.g., head shakes), and body contacts (e.g., holding, pushing) (*SI Appendix*). Coding of verbal behaviors included rule protest (i.e., explicitly referring to the sorting rule or using deontic verbs like "should") and imperative protest (e.g., "no," "don't do it"). Our statistical analyses focused on these codes (additional verbal codes are in *SI Appendix*). We analyzed our data with Bayesian multilevel models and compared sets of nested models using the widely application information criterion (WAIC) and WAIC weights (*Materials and Methods* and *SI Appendix*). We report estimates of fixed effects as incidence rate ratios, their 95% confidence intervals (95% CI), and posterior probabilities ($post.\ prob.$) for the direction of the effect (increase vs. decrease).

Comparing Observer Interventions across Conditions. We first analyzed whether observers intervened more frequently in the rule-conflict than in the same-rule condition. Model comparisons revealed that a model including condition, a condition × society interaction (modeled as random slopes for condition, varying by society), dyad age, and gender made better predictions ($WAIC = 1{,}963$, $SE = 47$, $weight = 1.0$) than models without the society × condition interaction or without main effects of society and condition (all other models: $WAICs \geq 2{,}022$, $SEs \geq 47$, $weights = 0$) (*SI Appendix*). We found strong evidence that observers were (255%) more likely to intervene in the rule-conflict condition as compared with the same-rule condition ($Est = 3.55$, 95% CI $= [2.32, 5.39]$, $post.\ prob.\ increase = 1.00$). We found evidence, but not definitive evidence, that older children were somewhat less likely to intervene than younger children[‡] ($Est = 0.91$, 95% CI $= [0.79, 1.03]$, $post.\ prob.\ decrease = 0.94$). The estimated effect of gender (boys vs. girls) was small and had an uncertain direction ($Est = 1.08$, 95% CI $= [0.83, 1.40]$, $post.\ prob.\ increase = 0.72$). To inspect the interaction between society and condition, we plotted expected means for each condition per society (Fig. 2). In all societies, expected mean interventions were higher in the rule-conflict condition than in the same-rule condition, and in all societies, apart from Leipzig, the 95% CIs did not overlap. In the rule-conflict condition, the model predicted the highest mean for Hai‖om children, the lowest mean for Leipzig children, and similar ranges for the other six societies. In supplementary analyses, we added the (natural logarithm of) community size (12) to the model, but the advantage of a model with an interaction between community size × condition was very small and uncertain as all models had very similar WAICs (*SI Appendix*). Estimates indicated (although not definitely) that the likelihood of intervention decreased slightly with increasing community size ($Est = 0.94$, 95% CI $= [0.86, 1.02]$, $post.\ prob.\ decrease = 0.95$). Additional analyses of observer and player behaviors can also be found in *SI Appendix*.

Comparing Verbal and Nonverbal Interventions (Rule-Conflict Condition). Next, we compared observers' use of verbal protest and nonverbal interventions in the rule-conflict condition. Model comparisons revealed that a model including intervention type,

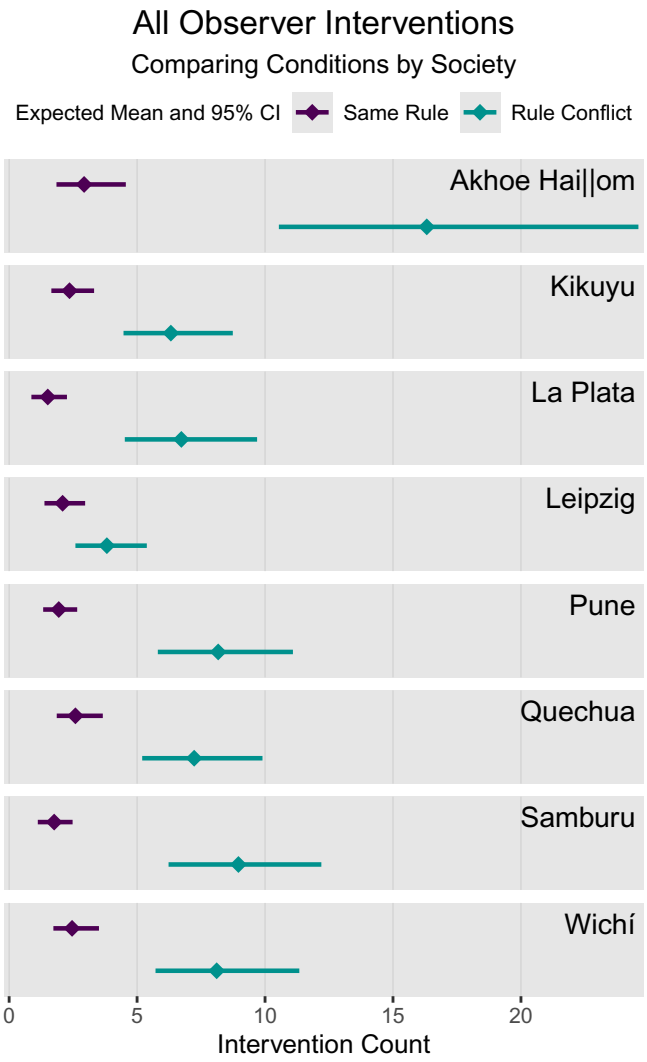

Fig. 2. All observer interventions (verbal protest and nonverbal interventions combined) across conditions. Expected means and 95% CIs were predicted for an average child per condition and society.

a type × society interaction (modeled as random slopes for type, varying by society), dyad age, and gender made better predictions ($WAIC = 1{,}745$, $SE = 35$, $weight = 1.0$) than any of the other models ($WAICs \geq 1{,}832$, $SEs \geq 44$, $weights = 0$) (*SI Appendix*). We find strong, but not definitive, evidence that children were more likely to use nonverbal interventions than verbal protest ($Est = 1.31$, 95% CI $= [0.92, 1.85]$, $post.\ prob.\ increase = 0.95$). Older children were less likely to intervene than younger children ($Est = 0.88$, 95% CI $= [0.78, 0.98]$, $post.\ prob.\ decrease = 0.99$). Gender had no clear effect ($Est = 1.00$, 95% CI $= [0.79, 1.27]$, $post.\ prob.\ increase = 0.49$). We plotted expected means for each society (Fig. 3) to further inspect the type × society interaction. Hai‖om children had clearly higher expected means for nonverbal than verbal interventions. For five societies (Kikuyu, La Plata, Quechua, Samburu, Wichí), the model predicted somewhat higher expected means for nonverbal interventions than for verbal protest, but 95% CIs overlapped considerably.

Comparing Types of Verbal Protest (Rule-Conflict Condition). We further analyzed observers' verbal protest by comparing their use of rule protest and imperative protest. Model comparisons revealed that a model including verbal type, a type × society interaction (modeled as random slopes for type, varying by society), dyad age, and gender made better predictions ($WAIC = 1{,}361$,

[‡]We *z*-transformed age so 1 unit corresponds to one SD. In our data, dyads had a mean age of *M* = 6.5 y with *SD* = 1.06 y.

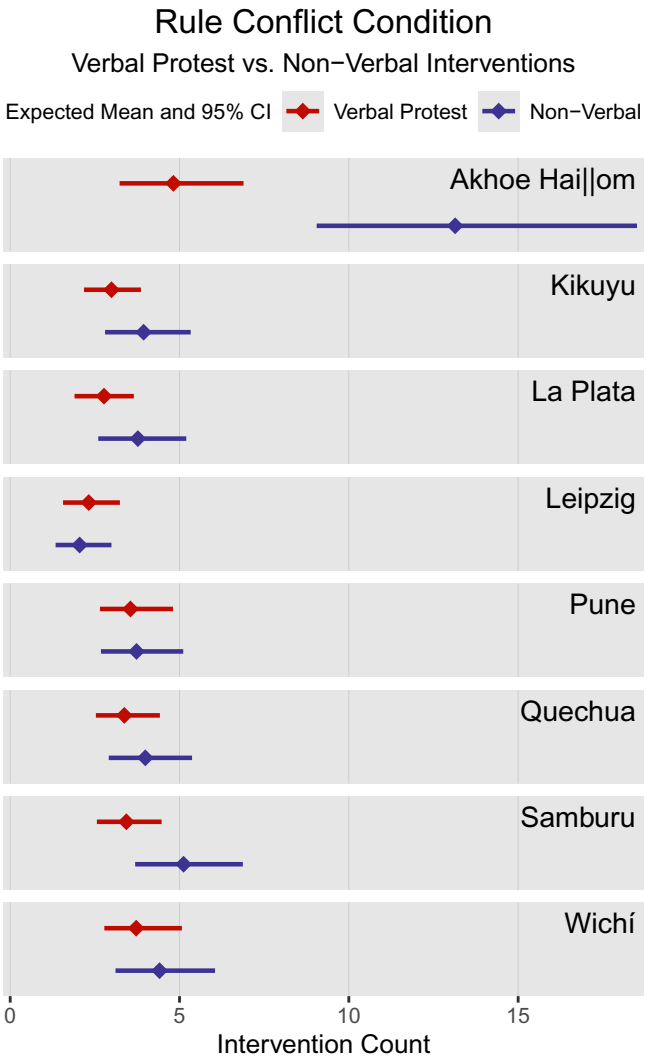

Fig. 3. Types of observer interventions (verbal, nonverbal) in the rule conflict condition. Expected means and 95% CIs were predicted for an average child per type and society.

$SE = 43$, $weight = 1.0$) than any of the other models ($WAICs \geq 1{,}428$, $SEs \geq 42$, $weights = 0$) (*SI Appendix*). We found strong evidence that children were more likely to use imperative protest than rule protest ($Est = 2.39$, 95% CI $= [1.23, 4.38]$, $post.\ prob.\ increase = 0.99$). There was no clear effect of age ($Est = 1.00$, 95% CI $= [0.89, 1.12]$, $post.\ prob.\ decrease = 0.50$). The estimated effect of gender was small and had an uncertain direction ($Est = 1.07$, 95% CI $= [0.86, 1.34]$, $post.\ prob.\ increase = 0.72$). Plotting of expected means for each society (Fig. 4) showed that in the five small-scale societies (Hai‖om, Kikuyu, Quechua, Samburu, Wichí), means were clearly higher for imperative protest than for rule protest (no overlap of 95% CIs). The three urban locations (La Plata, Leipzig, Pune) had similar expected means for the two verbal protest types, and their 95% CIs overlapped considerably. During verbal coding, we also scored whether children mentioned topics such as punishment, rewards, or authorities (e.g., experimenter/teacher) or tattled to the experimenter (*SI Appendix*). These categories were too infrequent for statistical analyses (descriptive statistics are in *SI Appendix*).

Comparing Types of Nonverbal Interventions (Rule-Conflict Condition). We also analyzed the different types of nonverbal interventions (actions, gestures, expressions, contacts). Model comparisons revealed that a model including nonverbal type, a type × society interaction (modeled as random slopes for type, varying by society), dyad age, and gender made better predictions ($WAIC = 2{,}190$, $SE = 88$, $weight = 1.0$) than any of the other models ($WAICs \geq 2{,}334$, $SEs \geq 96$, $weights = 0$) (*SI Appendix*). Across societies, children were more likely to use actions, gestures, and expressions than contacts (actions vs. contacts: $Est = 30.13$, 95% CI $= [12.61, 69.64]$; gestures vs. contacts: $Est = 30.27$, 95% CI $= [16.17, 59.74]$; expressions vs. contacts: $Est = 18.34$, 95% CI $= [8.65, 38.17]$; all $post.\ prob.\ of\ increase = 1.00$). Older children were less likely to use nonverbal interventions than younger children ($Est = 0.79$, 95% CI $= [0.69, 0.91]$, $post.\ prob.\ decrease = 1.00$). Gender had no clear effect ($Est = 0.94$, 95% CI $= [0.72, 1.22]$, $post.\ prob.\ of\ decrease = 0.69$). Plotting of expected means for the different intervention types revealed different societal profiles (Fig. 5) and no clear distinction between small-scale and large-scale societies.

Do Observer Interventions Affect Players' Sorting Rule?. Finally, we analyzed whether observers' overall interventions (i.e., verbal and nonverbal combined) predicted whether players switched to using their observer's sorting rule. Across societies, 20 of 188 players (10.6%) changed to the observer's sorting rule. Observers protested more often in dyads where the player adopted the observer's rule ($Mean = 22.3$, $SD = 10.4$) as compared with dyads where no change occurred ($Mean = 8.8$,

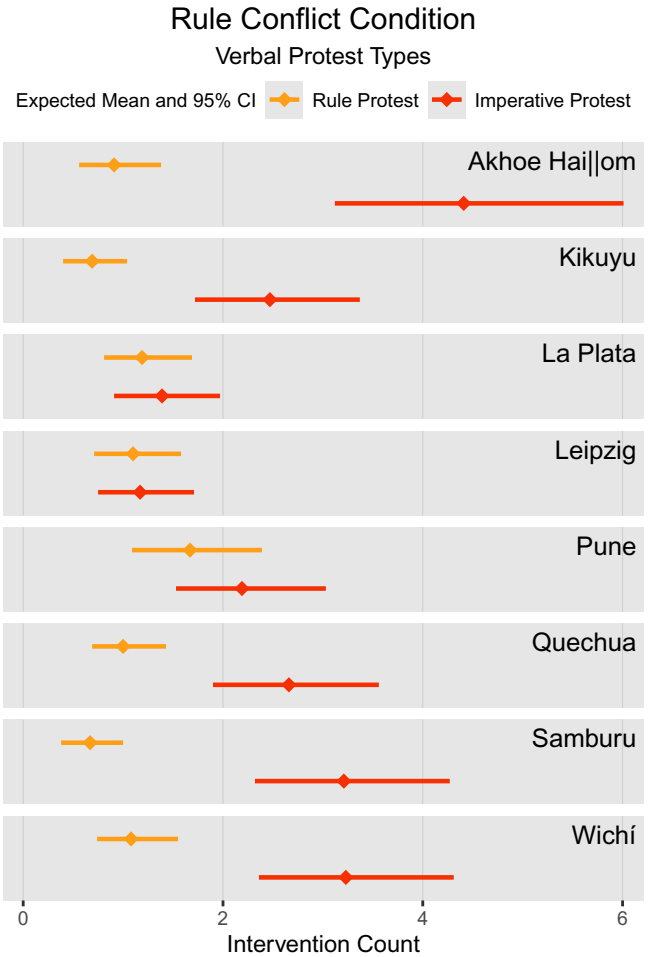

Fig. 4. Types of observers' verbal protest (rule protest, imperative protest) in the rule conflict condition. Expected means and 95% CIs were predicted for an average child per verbal type and society.

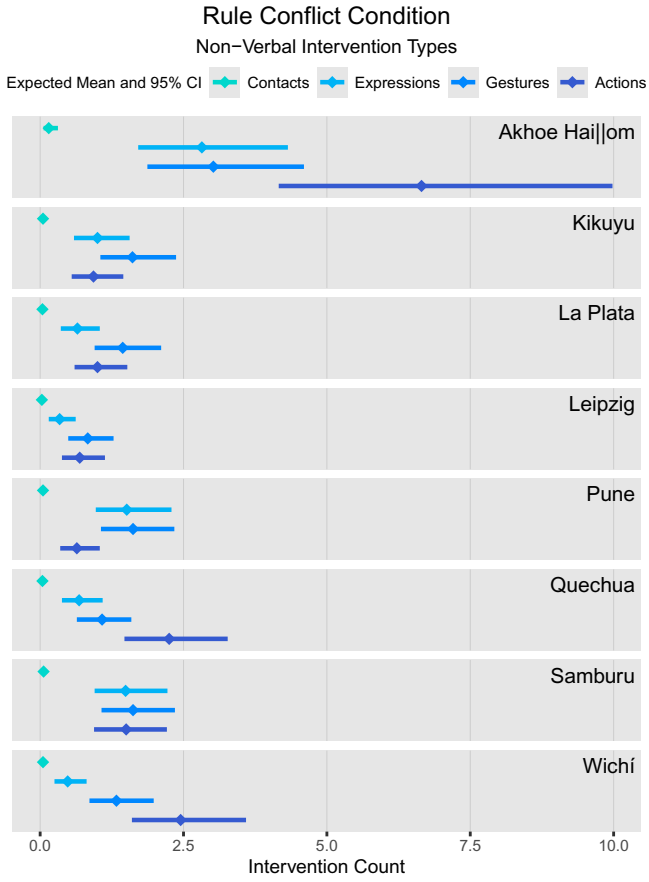

Fig. 5. Types of observers' nonverbal interventions (contacts, expressions, gestures, actions) in the rule conflict condition. Expected means and 95% CIs were predicted for an average child per nonverbal type and society.

$SD = 7.8$). Model comparisons revealed that a model including intervention frequency, society (random effect), dyad age, and gender ($WAIC = 106$, $SE = 18$, $weight = 0.31$) and a model without the effect of society ($WAIC = 104$, $SE = 18$, $weight = 0.69$) made better predictions than a model with only gender and age ($WAIC = 133$, $SE = 19$, $weight = 0$) (*SI Appendix*). The model without the random effect of society had a slight advantage, but WAICs for both models were very similar, so there remains uncertainty about which model better predicted the data. We report estimates for the more parsimonious model with higher relative support (note that estimates in the full model are almost identical); higher intervention rates increased the likelihood of a change in sorting rule ($Est = 1.15$, 95% CI $= [1.09, 1.21]$, $post.\ prob.\ increase = 1.00$), with a 15% (95% CI $= [9\%, 21\%]$) increase in likelihood per additional intervention. We found evidence (although not definitive) of a gender effect, with boys' interventions being more likely to result in change than girls' interventions ($Est = 2.39$, 95% CI $= [0.77, 7.53]$, $post.\ prob.\ increase = 0.93$). Age had no clear effect ($Est = 0.99$, 95% CI $= [0.58, 1.76]$, $post.\ prob.\ decrease = 0.52$).

General Discussion

Our study has four main findings. 1) Across eight highly diverse societies, children intervened more often to correct a peer who apparently broke the rules than a peer who followed the rules. 2) The intervention rates varied across societies, but there was no evidence that interventions increased with community size. 3) Verbal protest styles showed distinct patterns for small-scale vs. large-scale societies, but physical protest styles were more heterogeneous. 4) Observer interventions predicted whether players switched to the observer's rule.

Based on previous research and theoretical accounts, we expected lower third-party intervention rates in small-scale societies as compared with large-scale, urban societies (12, 14, 36) and speculated that hunter-gatherer groups may show the smallest effect due to their egalitarian social structure and high tolerance of children's noncompliance (37, 48). However, we found that Hai‖om children displayed the largest intervention effect (i.e., difference between same-rule and rule-conflict conditions). It is possible that the peer context of our study closely mirrored Hai‖om peer-play activities, which are largely unsupervised by adults and often involve rough play. The Wichí are the only other society in our sample with a hunter-gatherer past (although they have been sedentized for more than 100 y and are more market integrated than the Hai‖om), and they showed a similar intervention effect to five other (small- and large-scale) populations in our study. Leipzig children intervened less often than children from any of the other societies—including children from the other two urban locations (Pune, India; La Plata, Argentina). This aligns with recent findings that German adults, as compared with Indian and Argentinian adults, viewed it as less appropriate to use physical and verbal confrontation when others transgress norms and as more appropriate to use nonaction or gossip (20). Taken together, we found no evidence that children from large-scale societies intervened more frequently in conventional norm transgressions than children from small-scale societies. Future studies could investigate how cultural values (e.g., individual autonomy) as well as socioeconomic factors (e.g., median income) could explain variation across both small- and large-scale societies (20, 35).

Our results revealed variation in children's verbal and nonverbal intervention styles. We expected that children in urban, large-scale societies would use rule protest (e.g., "You should put red on red.") more frequently than imperative protest (e.g., "No, don't do it") (28, 29, 32), but we found that they used rule protest and imperative protest at similar levels. Previous studies employed puppets as norm violators, and it is possible that, in interactions with peers in our study, children assumed a shared understanding of the rules (common ground) and hence, saw less of a need to explicitly refer to the rules or to use normative language (49). Children in the five small-scale societies (Hai‖om, Kikuyu, Quechua, Samburu, Wichí) clearly preferred imperative protest over rule-related protest, and this variation in verbal protest styles between small-scale and large-scale societies could be potentially due to different cultural emphases on justification and deliberation (50). Note, however, that nonverbal interventions types were more dispersed and not uniform among small-scale and large-scale societies in our study. Societal variation in enforcement styles (or so-called metanorms) remains an understudied dimension of norm enforcement (20); yet, this variation matters as conflicts may arise not only about what norm to follow but also, about how to respond to norm transgressions.

One recent study showed that selective, costly third-party punishment of selfishness during resource distribution increased during middle childhood and could be detected reliably by age 9 to 10 y in four of six societies (13). We found that even younger children reliably enforced conventional norms as third parties when interventions had little to no costs. This is in line with previous findings for German and North American children who used third-party punishment at a younger age in cost-free or low-cost situations (51, 52). Our models also revealed for all intervention analyses, apart from verbal protest types, that older children were slightly less likely to intervene than younger children. As we used a spontaneous enforcement measure, this finding could be due to a number of nonmutually exclusive reasons. 1) Older children could have been more measured and controlled in their responses than younger children (53), 2) they could have reached

agreement faster than younger children (26), or 3) they may have tolerated conventional transgressions more because they recognized that different rules can coexist (54). Future studies could map norm enforcement across a wider age range (refs. 38 and 39 use such an approach) to gain insight into how third-party enforcement of conventional norms develops beyond early and middle childhood and for instance, collect concurrent measures on inhibitory control or on children's flexibility in reasoning about social rules.

Finally, there are a number of methodological points to consider. 1) We used a new sorting game with two incompatible conventional norms (i.e., sort by color vs. sort by shape). Game rules, like other conventional norms such as what side of the road to drive on or how to greet each other, are arbitrary rules about how we do things in a specific context (28, 29). They are particularly suited for experimental, cross-cultural, and developmental studies as there exists a potentially indefinite number of new conventional game rules, which allows, for example, for repeated studies in the same field site. Moreover, introducing new conventional rules avoids any potential negative consequences and ethical pitfalls of inducing children to violate established conventional norms of their cultural group. 2) We used a fixed-order design (the justification is in *SI Appendix*), and while we cannot entirely rule out that children may have become more comfortable with intervening later in the study, we are confident that interventions in the rule-conflict condition occurred primarily due to the apparent rule violation. Previous studies using both between- and within-subjects designs have consistently found that children protested more in rule-conflict conditions as compared with (no conflict) control conditions (26, 28–30). 3) Our study focused on children's spontaneous behavioral and verbal responses, but similar responses can be accompanied by different motivations. For instance, children may want to help a peer to avoid punishment or to teach them the correct rules (44). We coded children's utterances for references to topics such as punishment, reward, or (lack of) knowledge, but occurrences were too infrequent for any meaningful statistical analyses.

To summarize, we show that children across the globe enforce conventional norms, albeit in culturally variable ways. Conventional norms set our species apart from other animals (55) and are at the heart of uniquely human cultural diversity.

Materials and Methods

Participants. We analyzed data from 376 children (5 to 8 y old) from eight societies (Table 1). We chose this age range as piloting in some of the rural sites had revealed that younger children were often too shy to participate. Further details on the samples are in *SI Appendix*.

Ethics and Consent. The study was conducted in accordance with the Declaration of Helsinki and the ethical guidelines of the German Psychological Society and the Association of German Professional Psychologists. The study did not involve any invasive techniques, ethically problematic procedures, or deception, and therefore, it did not require approval by an institutional review board (the regulations on freedom of research are in the German Constitution, Section 5 [3]). Depending on locality, we obtained permission to conduct the research from education or science ministries, local school boards, and/or community representatives. Informed consent for children's study participation was given by head teachers (ǂAkhoe Haiǁom, Kikuyu, Quechua, Samburu, Wichí) or parents/legal guardians (La Plata, Leipzig, Pune).

Procedure. The sorting game consisted of a wooden box with two vertical sticks and 16 wooden blocks (further details are in *SI Appendix*). Blocks had two different shapes (round, square) and two different colors (red, blue). Children first participated in a training session and then, on a separate day, in the test session.

Training session. Children learned the sorting rules with a peer partner (not matched by gender or age). Half of the dyads learned to sort by color, and half learned to sort by shape. Children who participated as partners in the training session were never partnered in the test session. Further details on training are in *SI Appendix*.

Test session. In the test session, we combined children into sets of four matched by gender and age, allowing for up to a 1-y age gap (because of this, we use mean dyad age in all analyses). Each set included two children who had learned to sort by shape and two children who had learned to sort by color. We divided the test session into two phases (in fixed order): the same-rule phase and the rule-conflict phase (further justifications of the experimental design are in *SI Appendix*).

In the same-rule phase, we partnered children with a child who had learned the same sorting rule (color:color or shape:shape). At the start, children in a dyad played the game together to ensure that children 1) remembered the rules correctly and 2) viewed the rules as shared rules. After that, each child in a dyad once acted as an observer and once as a player (Fig. 1). We repeated this procedure for the second dyad in the set.

In the rule-conflict phase, we partnered children with a peer who—unknown to either child—had learned a different sorting rule (color:shape or shape:color). One child acted as an observer, and the other child acted as a player (Fig. 1), with assignment to roles determined by a fixed counterbalancing scheme (*SI Appendix* has details). There was only one trial and no role reversal as the rule conflict became obvious after the first trial. After that, children played the game together (not analyzed in the current study). We repeated this procedure for the second dyad in the set. At the end of the study, children received small gifts (e.g., sweets, stickers).

Data Recording and Coding. We recorded the study with two cameras for detailed analyses of children's behaviors and utterances. Children's utterances were transcribed and where necessary, translated by native speakers into English, German, or Spanish (Spanish was further translated into German). A single coder (fluent in German, English, and Spanish) scored children's nonverbal and verbal behaviors in all eight societies (*SI Appendix*). A second coder (also fluent in German, English, and Spanish) coded between 25 and 36% of data for reliability purposes. Reliability was very good, with the vast majority of behavioral codes $ICC3 \geq 0.79$ and verbal codes $\kappa \geq 0.78$ (*SI Appendix*).

Data analyses. We analyzed all data in R (56) and implemented Bayesian multilevel models using the brms package (57, 58) and the package's default priors. The majority of models were fitted using Poisson distributions, with the exception of the sorting change models that used a Bernoulli distribution. Our analytical strategy consisted of comparing sets of nested models that systematically varied in predictors of interest. The full model usually included an interaction of society and condition (or intervention type), specified as a random effect of condition/type, varying by society (a further explanation of this approach is in *SI Appendix*), and condition as a fixed effect. We also added mean dyad age (*z*-standardized to mean of zero and SD of one) and gender (coded as $-0.5/+0.5$ [female/male] so the intercept corresponds to the average of both genders) as fixed effects to the full model, as well as random effects of participant identifications (IDs). The null model only included age, gender, and random effects of participant IDs. Intermediate models had no interaction term (red1) or no random effect of society (red2). For example, we specified the nested models comparing observer interventions across conditions as follows.

Full: response ∼ dyad.age + gender + condition + (condition|society) + (1|observer.id) + (1|player.id)

Red1: response ∼ dyad.age + gender + condition + (1|society) + (1|observer.id) + (1|player.id)

Red2: response ∼ dyad.age + gender + condition + (1|observer.id) + (1|player.id)

Null: response ∼ dyad.age + gender + (1|observer.id) + (1|player.id)

For each model, we inspected chain convergence and efficiency using R-hat values and effective sample sizes, and we visually inspected trace plots of the Markov chains (59). We also conducted graphical posterior predictive checks by comparing observed data with simulated draws from the posterior predictive distribution using the bayesplot package (60, 61). To compare nested models, we used the WAIC, WAIC SEs, and WAIC weights (62). WAIC provides an approximation for the out-of-sample predictive accuracy of a model, with smaller WAIC values indicating better predictive accuracy. WAIC weights indicate relative support for a model within a set of compared models and always sum to one, with a higher weight indicating better relative support. For the model with the best support, we report estimates of fixed effects as incidence rate ratios, their 95% CIs, and posterior probabilities of the direction of the effect. We derived these posterior probabilities by calculating the ratio of estimates more than one

(i.e., increase) or less than one (i.e., decrease) for the respective fixed effect in the posterior sample. We also inspected posterior distribution plots (*SI Appendix*). To inspect random effects of condition|society or type|society, we predicted mean responses for an average child for each condition and society using the fitted function in brms and plotted expected means and their 95% CIs. Rcode to reproduce analyses and figures and deidentified data are at https://osf.io/x7r2q.

Data Availability. Anonymized behavioral data and reliability coding have been deposited in the Open Science Framework (https://osf.io/x7r2q; ref. 63). All shared data appears in the main text or *SI Appendix*.

ACKNOWLEDGMENTS. We thank the communities, schools, parents, and children for their support and for taking part in the study. We thank Carlos Diaz, Patricia Gutierrez, Shantanu Jadhav, Lucas Lepuyapui, Ewan Lesowopir, Richard Lesowapir, Gabriel Marzocchini, David Mundia Maina, Liza Muthoni Mundia, Julia Ohlendorf, Dennis Salazar Gonzales, Jahnavi Sunderarajan, and Disney Tjizao for their help with transcriptions and translations; Ronja Büchner, Laura Huber, Solveig Jurkat, Julia Ohlendorf, and Noelia Pereyra for helping with data collection; Elena Schmidt Velasco and Kathrin Kern for helping with data coding; Luke Maurits for statistical advise; and Cristina Zickert for drawing the study setup. This study was supported by a Freigeist Fellowship from Volkswagen Foundation Grant 48952 (to P.K.).

1. P. L. Berger, T. Luckmann, *The Social Construction of Reality: A Treatise in the Sociology of Knowledge* (Anchor Books, Garden City, NY, 1966).
2. M. Chudek, J. Henrich, Culture-gene coevolution, norm-psychology and the emergence of human prosociality. *Trends Cogn. Sci.* **15**, 218–226 (2011).
3. E. Fehr, U. Fischbacher, Third-party punishment and social norms. *Evol. Hum. Behav.* **25**, 63–87 (2004).
4. C. Bicchieri, *The Grammar of Society: The Nature and Dynamics of Social Norms* (Cambridge University Press, New York, NY, 2006).
5. R. Boyd, H. Gintis, S. Bowles, Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science* **328**, 617–620 (2010).
6. E. Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge University Press, Cambridge, United Kingdom, 1990).
7. E. Fehr, U. Fischbacher, Social norms and human cooperation. *Trends Cogn. Sci.* **8**, 185–190 (2004).
8. E. Turiel, "The development of morality" in *Social, Emotional, and Personality Development*, W. Damon, R. M. Lerner, N. Eisenberg, Eds. (Handbook of Child Psychology, Wiley, New York, NY, ed. 6, 2006), **vol. 3**, pp. 789–857.
9. E. Fehr, S. Gächter, Altruistic punishment in humans. *Nature* **415**, 137–140 (2002).
10. B. Herrmann, C. Thöni, S. Gächter, Antisocial punishment across societies. *Science* **319**, 1362–1367 (2008).
11. J. Henrich *et al.*, Costly punishment across human societies. *Science* **312**, 1767–1770 (2006).
12. J. Henrich *et al.*, Markets, religion, community size, and the evolution of fairness and punishment. *Science* **327**, 1480–1484 (2010).
13. B. R. House *et al.*, Social norms and cultural diversity in the development of third-party punishment. *Proc. Biol. Sci.* **287**, 20192794 (2020).
14. F. W. Marlowe, J. C. Berbesque, More 'altruistic' punishment in larger societies. *Proc. Biol. Sci.* **275**, 587–590 (2008).
15. S. Mathew, R. Boyd, Punishment sustains large-scale cooperation in prestate warfare. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 11375–11380 (2011).
16. S. Mathew, R. Boyd, The cost of cowardice: Punitive sentiments towards free riders in Turkana raids. *Evol. Hum. Behav.* **35**, 58–64 (2014).
17. P. Wiessner, The role of third parties in norm enforcement in customary courts among the Enga of Papua New Guinea. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 32320–32328 (2020).
18. H. C. Barrett *et al.*, Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4688–4693 (2016).
19. S. Nichols, *Sentimental Rules: On the Natural Foundations of Moral Judgment* (Oxford University Press, 2004).
20. K. Eriksson *et al.*, Perceptions of the appropriate response to norm violation in 57 societies. *Nat. Commun.* **12**, 1481 (2021).
21. D. Lewis, *Convention: A Philosophical Study* (Blackwell, Oxford, United Kingdom, 2008).
22. A. Dahl, T. Waltzer, Constraints on conventions: Resolving two puzzles of conventionality. *Cognition* **196**, 104152 (2020).
23. C. Bicchieri, R. Muldoon, A. Sontuoso, "Social norms" in *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, Ed. (Metaphysics Research Lab, Stanford University, 2018).
24. A. Gampe, M. M. Daum, How preschoolers react to norm violations is associated with culture. *J. Exp. Child Psychol.* **165**, 135–147 (2018).
25. S. Hardecker, M. F. H. Schmidt, M. Roden, M. Tomasello, Young children's behavioral and emotional responses to different social norm violations. *J. Exp. Child Psychol.* **150**, 364–379 (2016).
26. B. Köymen *et al.*, Children's norm enforcement in their interactions with peers. *Child Dev.* **85**, 1108–1122 (2014).
27. L. P. Nucci, M. S. Nucci, Children's social interactions in the context of moral and conventional transgressions. *Child Dev.* **53**, 403–412 (1982).
28. H. Rakoczy, F. Warneken, M. Tomasello, The sources of normativity: Young children's awareness of the normative structure of games. *Dev. Psychol.* **44**, 875–881 (2008).
29. H. Rakoczy, N. Brosche, F. Warneken, M. Tomasello, Young children's understanding of the context-relativity of normative rules in conventional games. *Br. J. Dev. Psychol.* **27**, 445–456 (2009).
30. M. F. H. Schmidt, L. P. Butler, J. Heinz, M. Tomasello, Young children see a single action and infer a social norm. *Psychol. Sci.* **27**, 1360–1370 (2016).
31. M. F. H. Schmidt, H. Rakoczy, M. Tomasello, Eighteen–month–old infants correct non–conforming actions by others. *Infancy* **24**, 613–635 (2019).
32. B. Tunçgenç, A. Hohenberger, H. Rakoczy, Early understanding of normativity and freedom to act in Turkish toddlers. *J. Cogn. Dev.* **16**, 44–54 (2015).
33. M. Nisan, Moral norms and social conventions: A cross-cultural comparison. *Dev. Psychol.* **23**, 719–725 (1987).
34. R. A. Shweder, M. Mahapatra, J. G. Miller, "Culture and moral development" in *The Emergence of Morality in Young Children*, J. Kagan, S. Lamb, Eds. (University of Chicago Press, Chicago, IL, 1987), pp. 1–83.
35. M. J. Gelfand *et al.*, Differences between tight and loose cultures: A 33-nation study. *Science* **332**, 1100–1104 (2011).
36. L. Glowacki, S. Lew-Levy, How small-scale societies achieve large-scale cooperation. *Curr. Opin. Psychol.* **44**, 44–48 (2021).
37. S. Lew-Levy, N. Lavi, R. Reckin, J. Cristóbal-Azkarate, K. Ellis-Davies, How do hunter-gatherer children learn social and gender norms? A meta-ethnographic review. *Cross Cult. Res.* **52**, 213–255 (2018).
38. B. R. House *et al.*, Ontogeny of prosocial behavior across diverse societies. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 14586–14591 (2013).
39. B. R. House *et al.*, Universal norm psychology leads to societal diversity in prosocial behaviour and development. *Nat. Hum. Behav.* **4**, 36–44 (2020).
40. P. R. Blake *et al.*, The ontogeny of fairness in seven societies. *Nature* **528**, 258–261 (2015).
41. H. C. Barrett, Towards a cognitive science of the human: Cross-cultural approaches and their urgency. *Trends Cogn. Sci.* **24**, 620–638 (2020).
42. P. Kanngiesser, F. Rossano, H. Zeidler, D. Haun, M. Tomasello, Children's respect for ownership across diverse societies. *Dev. Psychol.* **55**, 2286–2298 (2019).
43. M. Schäfer, D. B. M. Haun, M. Tomasello, Fair is not fair everywhere. *Psychol. Sci.* **26**, 1252–1260 (2015).
44. B. Köymen, M. F. H. Schmidt, L. Rost, E. Lieven, M. Tomasello, Teaching versus enforcing game rules in preschoolers' peer interactions. *J. Exp. Child Psychol.* **135**, 93–101 (2015).
45. A. Orvell, E. Kross, S. A. Gelman, That's how "you" do it: Generic you expresses norms during early childhood. *J. Exp. Child Psychol.* **165**, 183–195 (2018).
46. L. Li, B. Britvan, M. Tomasello, Young children conform more to norms than to preferences. *PLoS One* **16**, e0251228 (2021).
47. P. Wiessner, Norm enforcement among the Ju/'hoansi Bushmen: A case of strong reciprocity? *Hum. Nat.* **16**, 115–145 (2005).
48. T. Widlok, *Living on Mangetti: 'Bushman' Autonomy and Namibian Independence* (Oxford University Press, Oxford, United Kingdom, 1999).
49. B. Köymen, M. Mammen, M. Tomasello, Preschoolers use common ground in their justificatory reasoning with peers. *Dev. Psychol.* **52**, 423–429 (2016).
50. H. Keller, *Cultures of Infancy* (Lawrence Erlbaum Associates, 2007).
51. K. McAuliffe, J. J. Jordan, F. Warneken, Costly third-party punishment in young children. *Cognition* **134**, 1–10 (2015).
52. K. Riedl, K. Jensen, J. Call, M. Tomasello, Restorative justice in children. *Curr. Biol.* **25**, 1731–1735 (2015).
53. K. Schirmbeck, N. Rao, C. Maehler, Similarities and differences across countries in the development of executive functions in children: A systematic review. *Infant Child Dev.* **29**, e2164 (2020).
54. M. F. H. Schmidt, I. Gonzalez-Cabrera, M. Tomasello, Children's developing metaethical judgments. *J. Exp. Child Psychol.* **164**, 163–177 (2017).
55. L. Schlingloff, R. Moore, "Do chimpanzees conform to cultural norms?" in *The Routledge Handbook of Philosophy of Animal Minds*, K. Andrews, J. Beck, Eds. (Routledge Handbooks in Philosophy, Routledge Taylor & Francis Group, New York, NY, 2018), pp. 381–389.
56. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2020).
57. P. C. Bürkner, brms: An R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28 (2017).
58. P. C. Bürkner, Advanced Bayesian multilevel modeling with the R package brms. *R J.* **10**, 395–411 (2018).
59. A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, P. C. Bürkner, Rank-normalization, folding, and localization: An improved r for assessing convergence of MCMC. *Bayesian Anal.* **1**, 1–28 (2021).
60. J. Gabry, D. Simpson, A. Vehtari, M. Betancourt, A. Gelman, Visualization in Bayesian workflow. *J. R. Stat. Soc. Ser. A Stat. Soc.* **182**, 389–402 (2019).
61. J Gabry, T Mahr, bayesplot: Plotting for Bayesian models. R package version 1.8.0. https://mc-stan.org/bayesplot/. Accessed 5 May 2021.
62. R. McElreath, *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (Chapman and Hall/CRC, ed. 2, 2020).
63. P. Kanngiesser *et al.*, Children across societies enforce conventional norms but in culturally variable ways. Open Science Framework. https://osf.io/x7r2q. Deposited 7 July 2021.