# The Moral, or the Story? Changing Children's Distributive Justice Preferences Through Social Communication

Joshua Rottman, Valerie Zizik, Kelly Minard,
Liane Young, Peter R. Blake, & Deborah Kelemen

## Online Supplementary Materials

*Note: All data and R code are available at https://osf.io/sf6xa.*

## Study 1

**Method**

***Exploratory questions***

To determine the degree to which the stories were considered to be fantastical, each participant was asked to rate how real they thought the story was on a scale from 1 (definitely make-believe) to 4 (definitely real).

Additional dependent measures were added when it began to become evident that most children's post-test performance was not being impacted by the intervention. These were administered after all other measures to prevent any possible carryover effects. One exploratory measure, the "Structurally Similar Transfer Task" ($N = 56$) invoked analogically relevant features of the storybook content. In this task, children were told about a town in which people disagreed about how to divide firewood or water, and they were asked whether these resources should be divided equally or based on merit (the full text of these scenarios is available upon request). We reasoned that, if difficulty in transferring moral lessons from the storybook to the post-test distribution task was driven by an inability to map the beavers' wood distribution onto the distribution of prizes, this task would prove to be easier for participants. Alternatively, if the

difficulty in transferring from the storybook to the distribution task was driven by the taxonomic distance between beavers and humans, this task would prove to be difficult.

A subset of participants was additionally given a second exploratory measure: the "Squirrel Target Transfer Task" ($N = 30$). This was similar to the pre- and post-test resource distribution tasks, with the exception that the recipients were squirrels rather than humans and the resources were nuts rather than stickers or temporary tattoos. (Unlike the Structurally Similar Transfer Task, which simply asked for children's verbal assessment of fair outcomes, the Squirrel Target Transfer Task involved a behavioral resource distribution.) We reasoned that, if difficulty in transferring moral lessons from the storybook to the primary distribution task was driven by the distance between a beaver society and a human situation, children should be better able to apply the story's message when deciding how to distribute nuts between two squirrels (one of whom was described to be three times more productive than the other), as this more closely resembles the distribution of wood amongst beavers.

**Results**

*Fairness preferences at pre-test*

Children had a slight, non-significant tendency to be equality-distributors (38/70, or 54.29%) rather than merit-distributors (32/70, or 45.71%) at pre-test, $p = .550$.

*Additional exploratory analyses*

Participants were conflicted as to whether they believed the stories to be real ($M = 2.41$, $SD = 1.05$). They tended to endorse the reality of these stories marginally more in the Reasoned Storybooks conditions ($M = 2.68$, $SD = 1.11$) than in the Emotional Storybooks conditions ($M = 2.18$, $SD = 0.95$), $t(59.63) = 1.96$, $p = .055$. Reality endorsement levels remained consistent between children who changed their beliefs and children who did not change their beliefs,

suggesting that the perceived reality of the storybooks did not impact children's propensity to learn from them, $t(18.12) = 0.02$, $p = .984$.

On the Structurally Similar Transfer Task, 43 of 56 participants (76.79%) expressed beliefs consistent with those in the storybooks. This is significantly different from chance (50%), $p < .001$, thus providing evidence of successful transfer on a task involving humans that paralleled the storybooks. However, there was no association between this task and performance on the primary resource distribution post-test task, as only 11 of the 43 participants (25.58%) who expressed principles consistently with the storybook on this exploratory judgment task also divided stickers or tattoos in the distributive format that the storybook endorsed. This finding indicates that the animal/human divide may not have specifically driven the failure of the storybooks to influence resource distribution in the post-test measure. Rather, participants endorsed the story's lesson when asked about the hypothetical application of fairness principles when the content between the story and the task could be easily mapped, even if they failed to translate that into their actual resource distribution behaviors in a structurally dissimilar task.

On the Squirrel Target Transfer Task, exactly half of the participants (15/30) distributed nuts in accordance with the lessons espoused by the storybook. Because this task had no pre-test measure, this result is exactly what would be predicted if children were choosing randomly, and therefore this cannot provide any evidence that children transferred the information from the storybooks to their performance on this task. Additionally, there was no association between performance on this task and performance on the post-test resource distribution task: only 3 of the 15 participants who divided nuts in the distributive format that the storybook endorsed (20.00%) also divided stickers/tattoos in the distributive format that the storybook endorsed.

In comparing results from the Structurally Similar Transfer Task and the Squirrel Target Transfer Task, it appears that the structure of a post-test scenario is more likely to contribute to children altering their beliefs about fairness as compared to the taxonomic closeness of recipients to characters in the stories. However, changing expressed beliefs does not seem to translate into changing distribution decisions.

### *Reanalysis of primary results with a sample of 64 (first 16 participants per condition)*

When excluding the six additional participants who were read the Emotional Merit Storybook, results remain nearly identical. Overall, 15/64 (23.44%) of participants changed their distributions after being read a storybook, which is very similar to the rate of change we found in the Baseline study (21.88%), as confirmed by a one-sample binomial test using this latter percentage as a point estimate, $p = .763$. Tendencies for changing distribution patterns did not differ from the Baseline study pattern in any of the four conditions, $p$s > .36. Logistic regressions indicated that changes in resource distribution did not differ across Fairness Type, Appeal, the two-way interaction between these variables, Age in Months, or Gender, $p$s > .41.

### *Justifications of resource distribution*

We analyzed participants' justifications to determine whether children tended to change the content of their reasoning in accordance with changes in resource distributions. Two independent coders coded children's justifications for their resource divisions. The coders demonstrated substantial agreement (Pre-test: $\kappa = .70$; Post-test: $\kappa = .83$). All disagreements were resolved through mutual discussion alongside the first author.

Overall, 14/16 (87.50%) of participants who changed their distribution patterns also changed the content of their justification for their distribution between pre-test and post-test. This pattern held true for only 16/54 (29.63%) of participants whose distributions remained consistent

between pre-test and post-test (see Fig. 3 in the main text). A logistic regression was conducted to predict changes in the content of justifications from Resource Distribution Change (No Change at Post-Test vs. Change at Post-Test), Fairness Type (Merit Intervention vs. Equality Intervention), and Appeal (Emotional Intervention vs. Reasoned Intervention). This analysis confirmed that participants who altered their distribution patterns from equality to merit or from merit to equality were more likely to change the content of their justification from pre-test to post-test than participants whose distribution patterns remained constant, $b = 3.60$, $p < .001$. There was also a difference across the two kinds of Appeal, $b = 1.71$, $p = .017$, such that emotional storybooks led to more change in justification types than reasoned storybooks. Finally, there was a marginally significant tendency for participants to change the content of their justifications across the two kinds of Fairness, $b = -1.13$, $p = .073$.

## Study 2

### Results

### *Fairness preferences at pre-test*

Participants were significantly more likely to be equality-distributors (69/101, or 68.32%) rather than merit-distributors (32/101, or 31.68%) at pre-test, as demonstrated by a one-sample binomial test, $p < .001$.

### *Reanalysis of primary results with a sample of 64 (first 16 per condition)*

Excluding the extra 37 children who were tested in the Merit Testimony conditions yields very similar results. A one-sample binomial test (comparing to a baseline rate of 21.88% change) demonstrated that testimony reliably led to changes in children's distribution patterns from pre-test to post-test, $p < .001$, with 68.75% of participants changing their distribution patterns after

testimony was presented. This remained true when examining each condition separately, all $ps <$ .003. Logistic regressions indicated that changes in resource distribution did not differ across Fairness Type, Appeal, the two-way interaction between these variables, Age in Months, or Gender, $ps > .13$ (but again, Fairness Type was a significant predictor when it was the only variable entered into the model, $b = 1.22$, $p = .035$).

### *Justifications of resource distribution*

Two independent coders coded children's justifications for their resource divisions. Each justification was assigned a single code based on which of five predefined categories seemed most representative (see Table 1). The two coders demonstrated substantial agreement (Pre-test: $\kappa = .75$; Post-test: $\kappa = .76$). All disagreements were resolved through mutual discussion alongside the first author.

Overall, 59/68 (86.76%) of participants who changed their distribution patterns also changed the content of their justification for their distribution between pre-test and post-test. This pattern held true for only 7/33 (21.21%) of participants whose distributions remained consistent between pre-test and post-test (see Fig. 3). A logistic regression was conducted to predict changes in the content of justifications from Resource Distribution Change (No Change at Post-Test vs. Change at Post-Test), Fairness Type (Merit Intervention vs. Equality Intervention), and Appeal (Emotional Intervention vs. Reasoned Intervention). This analysis confirmed that participants who altered their distribution patterns from equality to merit or from merit to equality were more likely to change the content of their justification from pre-test to post-test than participants whose distribution patterns remained consistent, $b = 3.39$, $p < .001$. There were no significant effects of Fairness Type or Appeal, $ps > .25$.

These results indicate that children who changed their patterns of resource distribution frequently shifted the content of their justifications. Children who became equality-distributors after being exposed to testimony advocating equality provided justifications that very closely resembled the justifications provided by children who preferred equality at pre-test (primarily focusing on considerations of recipients' welfare and moral principles). Approximately half of children who became merit-distributors after being exposed to testimony advocating merit provided justifications that resembled the justifications provided by children who preferred merit at post-test (primarily focusing on considerations of outputs). Intriguingly, the other half of these new merit-distributors' justifications focused on considerations of inputs (e.g., ability, effort), which were conspicuously absent during pre-test. This suggests that, while 6- to 7-year-olds do not tend to spontaneously justify merit-based distributions by appealing to differential inputs, interventions invoking these reasons seem to have a pronounced impact on children's decisions to allocate more resources to people who work harder as well as a pronounced impact on these children's subsequent reasoning for their allocations. More tentatively, it is possible that some children learned something new during the merit-based intervention, rather than switching their preferences to a latent form of fairness that was merely unexpressed during the pre-test.

## Study 3

All methods, exclusion criteria, and analyses for Study 3 were preregistered at https://osf.io/sf6xa. However, some analytic decisions were changed after collecting data in order to reduce the number of tests that were run (in particular, a logistic regression was conducted to test the predictors of participants' changes in distribution patterns, rather than a series of chi-

square tests). The overall pattern of results remains consistent when the preregistered analyses are conducted instead.

**Method**

*Exploratory questions*

At the end of the initial testing session, participants were asked whether or not the intervention was trying to teach them something. They were also asked to rate the realism of the storybook or video testimony on a scale from 1 (definitely make-believe) to 4 (definitely real).

After completing the delayed post-test distribution task, participants were asked when they learned to divide prizes the way they did, in order to assess the extent to which children were aware of the influence of the intervention. Additionally, participants were again asked the comprehension questions from the previous testing session.

*Parent questionnaire*

Parents ($N = 75$) completed a brief questionnaire during the initial session. First, the pre-test/post-test distribution task was described to them, and they were asked to indicate their own fairness preferences and to predict their children's fairness preferences. They also provided information about their children's sharing tendencies, information about their socialization practices at home, information about their social values, and demographic information.

**Results**

*Fairness preferences at pre-test and across timepoints*

Participants were non-significantly more likely to be equality-distributors (45/77, or 58.44%) than merit-distributors (32/77, or 41.56%) at pre-test, as indicated by a one-sample binomial test, $p = .171$. Overall, there were similar frequencies of merit-based and equality-based distributors at all three timepoints, $Q(2) = 1.16$, $p = .559$.

*Exploratory analyses*

Participants generally tended to believe that the stories and videos were real ($M = 3.21$, $SD = 0.82$). They endorsed the reality of the intervention more for Testimony ($M = 3.46$, $SD = 0.63$) than Storybooks ($M = 3.00$, $SD = 0.92$), $t(70.71) = 2.58$, $p = .012$. Reality endorsement levels were marginally higher for children who did *not* change their beliefs ($M = 3.37$, $SD = 0.80$) as compared to children who changed their beliefs ($M = 3.04$, $SD = 0.83$), $t(72.85) = 1.74$, $p = .086$, indicating that learning was not facilitated by perceiving interventions as realistic.

Participants typically reported that the intervention was trying to teach them something. This was more pronounced for participants who changed their fairness preferences (35/36, or 97.22%), as compared to those who did not (33/41, or 80.49%), $p = .032$.

Participants who demonstrated a changed fairness preference at the delayed post-test were likely to report that they learned how to divide resources during the previous session (25/40, or 62.50%), as compared to participants who did not demonstrate a changed fairness preference at the delayed post-test (3/28, or 10.71%), $b = 2.63$, $p < .001$.

*Reanalysis of primary results with a sample of 64 (first 16 per condition)*

Excluding the extra 13 children who were tested in the Merit conditions yields similar results. One-sample binomial tests comparing against baseline rates of change (21.88%) demonstrated that, collapsed across conditions, the interventions reliably led to changes in children's distribution patterns from pre-test to post-test, $p < .001$, with 50.00% change overall, and from pre-test to delayed post-test, $p < .001$, with 61.02% change. A significant change was found for both interventions preaching equality, both at the initial post-test and the delayed post-test, $ps < .001$. For the interventions preaching merit, there were not significant changes during

the immediate post-test (Merit Storybook: $p = .367$; Merit Testimony: $p = .061$), although there were at the delayed post-test (Merit Storybook: $p = .029$; Merit Testimony: $p < .001$).

A logistic regression, including only participants who returned for the delayed post-test, was conducted to predict whether children were differentially likely to change their style of resource distribution across the two Fairness Types (Merit Intervention vs. Equality Intervention), the two kinds of Intervention (Storybooks vs. Testimony), the Timepoint (Immediate vs. Delayed), and the interaction of these variables. Changes in resource distribution marginally differed across the two kinds of Fairness, $b = 1.42$, $p = .072$. Despite this difference becoming descriptively less pronounced at the delayed post-test, there were no significant effects of Timepoint, $b = 0.88$, $p = .260$, or of the interaction between Fairness Type and Timepoint, $b = -0.59$, $p = .588$. Changes in patterns of resource distribution did not differ across the two kinds of Intervention (Storybook vs. Testimony), $b = 1.01$, $p = .201$, and there were no significant two-way or three-way interactions involving this variable, $ps > .35$.

Adding Age in Months and Gender to the model weakened the effect of Fairness Type, $b = 0.91$, $p = .277$. The effect of Gender was insignificant, $b = -0.17$, $p = .668$, but Age predicted change, $b = 0.53$; $p = .041$. As there were no other significant effects in this model, we removed all predictors aside from Age. This led Age to become an even stronger predictor, $b = 0.63$, $p = .006$.

### *Justifications of resource distribution*

Two independent coders coded children's justifications for their resource divisions. Each justification was assigned a single code based on which of five predefined categories seemed most representative (see Table 1). The two coders demonstrated substantial agreement (Pre-test:

$\kappa$ = .82; Post-test: $\kappa$ = .81; Day 2: $\kappa$ =.74). All disagreements were resolved through mutual

discussion alongside the first author.

Overall, 35/36 (97.22%) of participants who changed their distribution patterns on Day 1

also changed the content of their justification for their distribution between pre-test and post-test

on Day 1. This pattern held true for only 6/41 (14.63%) of participants whose distributions

remained consistent between pre-test and post-test on Day 1 (see Fig. 3). A logistic regression

was conducted to predict changes in the content of justifications from Resource Distribution

Change (No Change at Post-Test vs. Change at Post-Test), Fairness Type (Merit Intervention vs.

Equality Intervention), and Version (Storybook vs. Testimony). This analysis confirmed that

participants who altered their distribution patterns from equality to merit or from merit to

equality were more likely to change the content of their justification from pre-test to post-test

than participants whose distribution patterns remained constant, $b$ = 5.41, $p$ < .001. There were

no significant effects of Fairness Type or Version, $p$s > .47.

### *Parent questionnaire*

Unlike children, parents had a slight preference for merit (60%) over equality (40%) on

the resource distribution task, $p$ = .105. There was no significant association between parents'

fairness preferences and their children's fairness preferences, $\chi^2(1)$ = 0.93, $p$ = .336.

Generally, responses on the parent questionnaire were not predictive of children's

fairness preferences, children's tendencies to change their distribution patterns, or parents'

fairness preferences ($p$s > .08). Parents were able to accurately predict their children's fairness

preferences at pre-test, $\chi^2(1)$ = 8.15, $p$ = .004, but because we did not closely monitor parents as

they completed the survey, it is possible that some parents discovered their children's pre-test

preferences before answering this question; as such, we are not confident in this result.