Supplemental Materials
"Preschoolers use the means-ends structure of intention to make moral judgments"

Included in this file:
1. Prescreening Materials and Methods
2. Materials and Methods for Studies 1 & 2
3. Supplemental Figures for Studies 1 & 2
4. Methods, Results and Figures for Study 2b

1. Prescreening Materials and Method

At the beginnings of all studies reported in this paper, subjects were trained to use the following
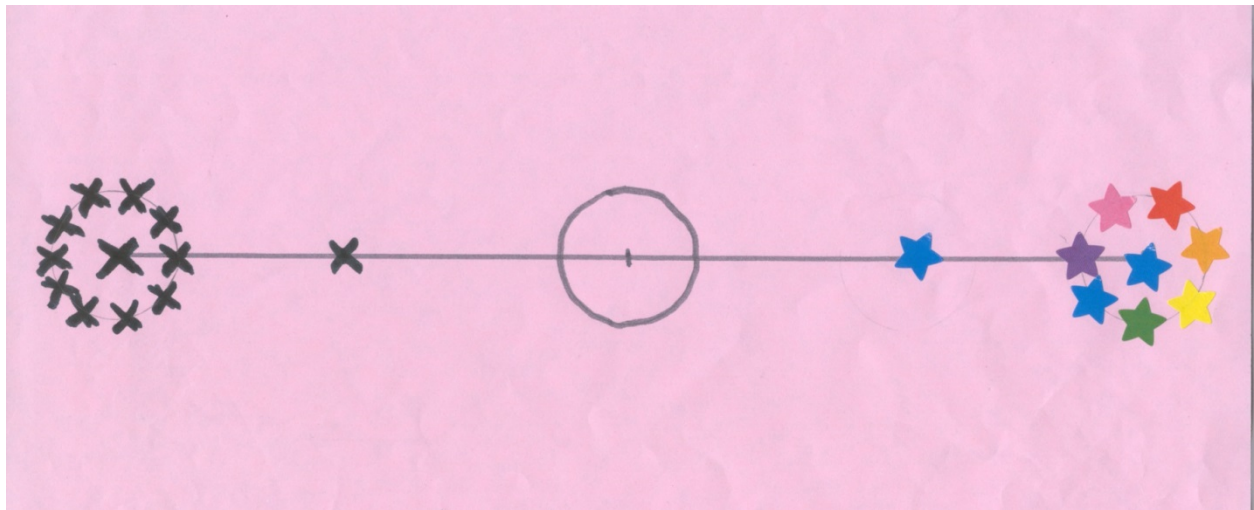
scale:

Fig. S1: Likert scale ("pink scale") used in all studies.

Children were given the following explanation of the scale:

This is called the Pink Scale Game and in this game we show each other when things are good [point to stars], bad [point to x's] or just ok [point to circle]. First let's think of something good. Can you think of something good? [Wait for child to respond.] Is that really good [point to lots of stars] or just a little good [point to one star]?

Then encourage child to offer a suggestion of something that is a little good or really good, whichever they haven't already offered.

If child can't think of something good at all (or can't think of something really good or a little good) offer a suggestion such as "eating an apple" or "helping your teacher" or "playing outside."

Repeat with "bad" and "just OK". Child should express competence using all five points on the scale (really good, a little good, just OK, a little bad, and really bad).

Children are then told two stories accompanied by pictures:

<u>Screening Story 1</u>

This is a story about Billy and Johnny. In this story, Billy hits Johnny.

Should Question: Should Billy have done that? (Inclusion Answer: No)

Action Rating: Can you show me on the pink scale? Was what Billy did: good, bad, or just OK?

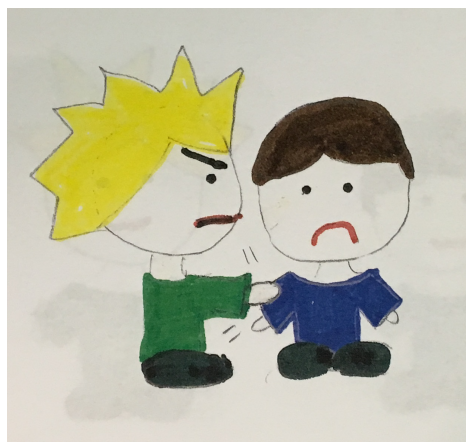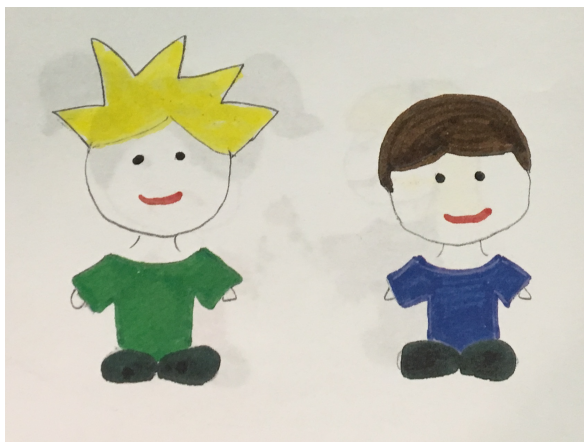(Inclusion Answer: Really bad or a little bad.)

Fig. S2: Images for screening story 1.

Screening Story 2

This is a story about Sue and Anne. What is Sue holding? That's right, a flower! In this story, Sue gives her flower to Anne.

Should Question: Should Sue have done that? (Inclusion Answer: Yes.)

Action Rating: Can you show me on the pink scale? Was what Sue did: good, bad, or just OK?

(Inclusion Answer: Really good, a little good, or just OK.)

Fig. S3: Images for screening story 2.

Note: To be included in the study, children needed to provide the inclusion answers to both the Should Question and Action Rating Question for both a good and a bad story. This was to ensure that children expressed competence answering the Should Question in both the negative and the affirmative and that they could use the Pink Scale competently. If children failed either question on the bad story (Story 1), they were given another bad story (Story 4, below); if they failed either question on the good story (Story 2), they were given another good story (Story 3, below). If they failed the second stories, they were excluded from the study for failing the screening.

Screening Story 3

This is a story about Billy and Johnny. In this story, Billy has a cookie and he gives it to Johnny.

Should Question: Should Billy have done that? (Inclusion Answer: Yes.)

Action Rating: Can you show me on the pink scale? Was what Billy did: good, bad, or just OK? (Inclusion Answer: Really good, a little good, or just OK.)
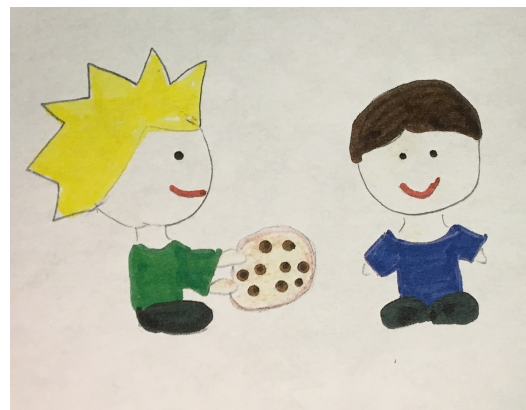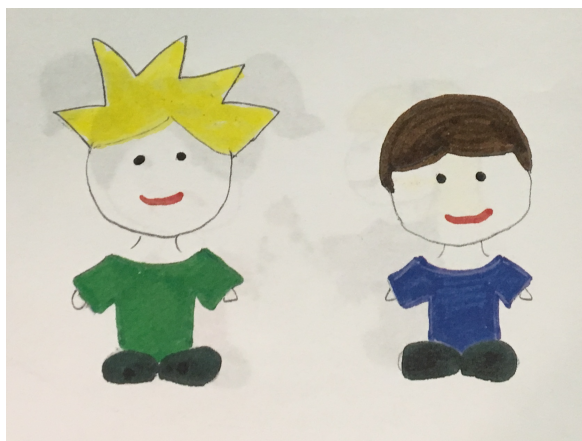

Fig. S4: Images for screening story 3.

This is a story about Sue and Anne. What is Sue holding? That's right, a flower! In this story, Anne takes Sue's flower and she breaks it.

Should Question: Should Anne have done that? (Inclusion Answer: No.)

Action Rating: Can you show me on the pink scale? Was what Anne did: good, bad, or just OK? (Inclusion Answer: Really bad or a little bad.)


Fig. S5: Images for screening story 4.

2. Materials and Methods for Studies 1 & 2

Note: Stimuli are available at https://github.com/sydneylevine/preschool-moral-intention

Each subject heard the story and the animation twice. On the first telling, subjects were corrected if they answered questions imbedded in the story incorrectly (eg, "What do you think the squirrel is going to do?"). On the second telling, subjects were asked a more extensive set of questions, ensuring that they understood key features of the scenario. Some of these questions acted as exclusion criteria (as indicated below).

Study 1: Materials and Methods

| Means | Side-Effect | Means+Stealing |
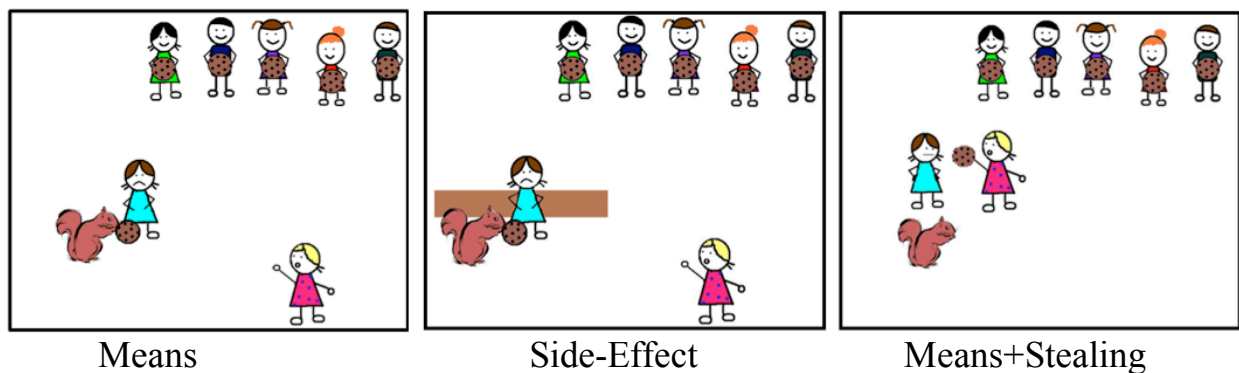
Selected scenes from the three conditions of Study 1.

Means Condition

This is a story about Katy. And Katy is playing in the park. There are some other kids in this story too. There is one kid over here and lots of kids over here. Today, all these kids are eating cookies! These kids all have cookies. But ut oh, here comes a mean, sneaky squirrel. He likes to eat other people's food. That squirrel is coming over here and -- what do you think he's going

to do? That's right, he wants to come over here and eat all these kids cookies. Not this kid [point to the one], just these kids [the five]. And how are these kids going to feel if the squirrel eats their cookies? That's right, they'll feel sad. So Katy knows what the squirrel is going to do. Katy knows that the squirrel is going to go eat those kids cookies and make them feel sad. So let's see what she does!

Katy waves to this girl over here. Katy waves to her and says, "Hey, come over here!" So that girl comes over. And Katy knows that when the girl gets to here, that she will get in the way of the squirrel. And the squirrel will eat her cookie. So that girl is sad because she doesn't get to eat her own cookie. But these kids are not sad, because they get to eat their own cookies.

Ok, let's watch that again from the beginning. [Replay the animation, advancing to the relevant screen with each control question.]

In the beginning, do you remember where the squirrel was going to go?

How were those kids going to feel?

Control questions (subjects were excluded for incorrect answers):

What did Katy do?

Where did the girl go?

What did the squirrel do?

How did that girl feel?

Did these kids feel sad?

So that's the end of the story, but I'm wondering about something. I'm wondering about Katy and what she did. In this story, Katy waved to the girl and told her to come over.

[Should Question] Should she have done that?

[Action Rating] Can you show me on the pink scale. Was what Katy did, good, bad, or just OK?

Side-Effect Condition

This is a story about Katy. And Katy is playing in the park. There are some other kids in this story too. There is one kid over here and lots of kids over here. This girl over here as a gate with her. She just carries around her gate with her just like that. Today, all these kids are eating cookies! These kids all have cookies. But ut oh, here comes a mean, sneaky squirrel. He likes to eat other people's food. That squirrel is coming over here and -- what do you think he's going to do? That's right, he wants to come over here and eat all these kids cookies. Not this kid [point to the one], just these kids [the five]. And how are these kids going to feel if the squirrel eats their cookies? That's right, they'll feel sad. So Katy knows what the squirrel is going to do. Katy knows that the squirrel is going to go eat those kids cookies and make them feel sad. So let's see what she does!

Katy waves to this girl over here. Katy waves to her and says, "Hey, bring your gate over here!" So that girl brings her gate over. Katy knows that when the girl gets here, that the squirrel will eat her cookie. Did Katy know that that would happen? And Katy also knows that when the girl gets to her, that her gate will block the squirrel. And the squirrel cannot get

passed. Did Katy know that that would happen? So that girl is sad because she doesn't get to eat her own cookie. But these kids are not sad, because they get to eat their own cookies.

Ok, let's watch that again from the beginning. [Replay the animation, advancing to the relevant screen with each control question.]

In the beginning, do you remember what this girl has with her?

Where was the squirrel was going to go?

How were those kids going to feel?

Control questions (subjects were excluded for incorrect answers):

What did Katy do?

Where did the girl go?

What did the squirrel do?

Did Katy know that that would happen?

How did that girl feel?

Did the squirrel get passed the gate?

Did Katy know that that would happen?

Did these kids feel sad?

Test questions:

So that's the end of the story, but I'm wondering about something. I'm wondering about Katy and what she did. In this story, Katy waved to the girl and told her to come over.

[Should Question] Should she have done that?

[Action Rating] Can you show me on the pink scale. Was what Katy did, good, bad, or just OK?

Means+Stealing Condition

This is a story about Katy. And Katy is playing in the park. There are some other kids in this story too. There is one kid over here and lots of kids over here. Today, all these kids are eating cookies! These kids all have cookies. But ut oh, here comes a mean, sneaky squirrel. He likes to eat other people's food. That squirrel is coming over here and -- what do you think he's going to do? That's right, he wants to come over here and eat all these kids cookies. Not this kid [point to the one], just these kids [the five]. And how are these kids going to feel if the squirrel eats their cookies? That's right, they'll feel sad. So Katy knows what the squirrel is going to do. Katy knows that the squirrel is going to go eat those kids cookies and make them feel sad. So let's see what she does!

Katy waves to this girl over here. Katy waves to her and says, "Hey, come over here!" And Katy knows that when the girl gets to here, that she is going to take her cookie and feed it to the squirrel. And the squirrel will eat her cookie. So that girl is sad because she doesn't get to eat her own cookie. But these kids are not sad, because they get to eat their own cookies.

Ok, let's watch that again from the beginning. [Replay the animation.]

In the beginning, do you remember where the squirrel was going to go?

How were those kids going to feel?

Control questions (subjects were excluded for incorrect answers):

What did Katy do?

Where did the girl go?

What did the squirrel do?

How did that girl feel?

Did these kids feel sad?

So that's the end of the story, but I'm wondering about something. I'm wondering about Katy and what she did. In this story, Katy took this girl's cookie and fed it to the squirrel. [Should Question] Should she have done that?

[Action Rating] Can you show me on the pink scale. Was what Katy did, good, bad, or just OK?

Note on Test Questions: If a child said "I don't know" or something irrelevant, we prompted them with the question again. For example, "What do you think? Should she have done that?" If subjects responded with two different answers at once, (for example, "yes and no" or "good and bad"), we would ask (for example), "If you had to choose one, what would you say? Should she have done that?" If a subject still failed to provide an answer, they were excluded from the analysis.

Study 1: Sample Size and Power Analysis

We chose our sample size based on sample sizes used in a previous unpublished study in our lab that used a similar paradigm (Saunders, 2014). However, some may be interested to note that a power analysis shows that these sample sizes were appropriate assuming a moderate effect size (0.4, based on the effect reported in Saunders, 2014), a desired power of 0.9, and an α-level of .05. We computed that a sample of 80 subjects was necessary (over the three conditions being analyzed) to find a difference between the conditions with a χ^2 analysis for subjects' Should

Question judgments (power analysis was done with G*Power 3, see Faul, Erdfelder, Lang, & Buchner, 2007).

Study 1: Supplemental Analysis

For this analysis, subjects in the Side-Effect condition were only excluded if they failed a control question that also appeared in the Means and Means+Stealing Condition, as explained in the main text. Ultimately, this meant that no subjects were excluded for failing control questions. Therefore, the Means and Means+Stealing Conditions contain 42 subjects each and the Side-Effect Condition contains 50 subjects.

In the Means Condition, 26.2% of subjects (11 out of 42) responded "yes" to the Should Question, $CI_{95\%}$= [13%, 39%]. In the Side-Effect Condition, 60.0% of subjects (30 out of 50) said "yes" to the Should Question, $CI_{95\%}$= [52%, 81%]. In the Means+Stealing Condition, 11.9% of subjects (5 out of 42) responded "yes" to the Should Question, $CI_{95\%}$= [2%, 22%].

There was a significant effect of condition on answers to the Should Question, $\chi^2_{Pearson}$ (2, n = 134) = 25.22, V_{Cramer} = 0.42, $CI_{95\%}$ [0.23, .58], p<.001, BF>100 in favor of H_1. Subjects were significantly more likely to answer "yes" to the Should Question in the Side-Effect Condition than in the Means Condition, χ^2 (1, n = 92) = 10.56, V_{Cramer} = 0.32, $CI_{95\%}$ [0.12, .53], p<.001. There was no significant difference between subjects' responses to the Means Condition and the Means+Stealing Condition, $\chi^2_{Pearson}$ (1, n = 84) = 2.78, V_{Cramer} = 0.18, $CI_{95\%}$ [-.02, .27], p=.095, BF=1.27 in favor of H_1 (Fig. S6).
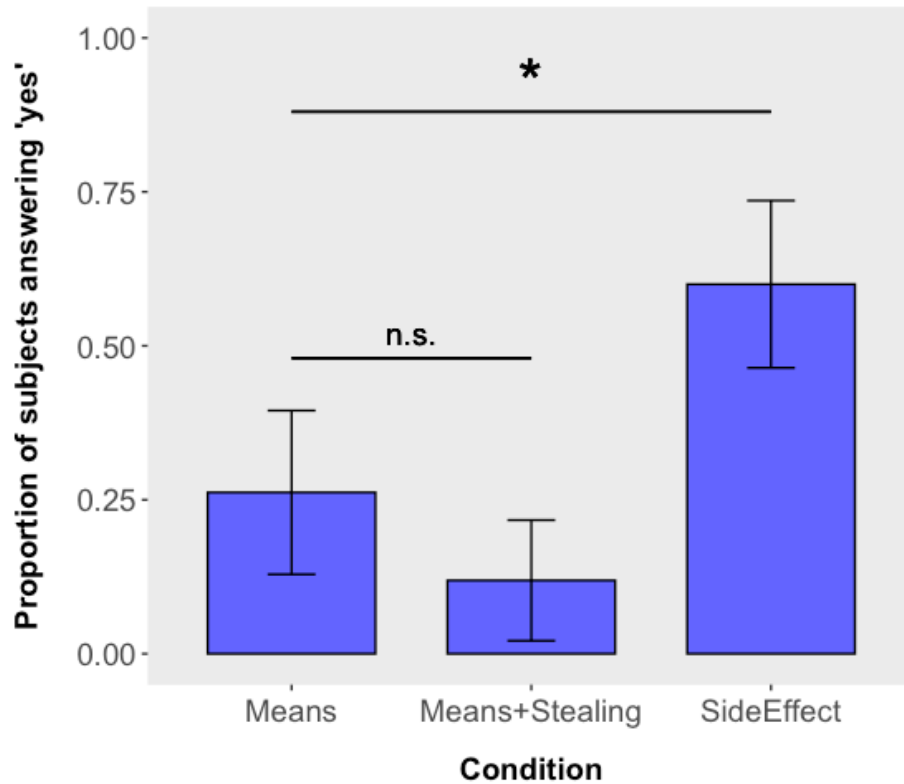
Fig. S6.

Proportion of subjects responding *yes* to the question, "Should she have done that?" as a function of condition. Error bars are 95% confidence intervals. The asterisks (*) indicates the significant difference between the Means and Side-Effect Conditions. There was no significant difference (n.s.) between the Means and Means+Stealing Conditions.

* $p < .001$

Likert scale ratings (here, and in all subsequent studies) were scored as follows: really bad = -2; a little bad = -1; just OK = 0; a little good = 1; really good = 2. Subjects in the Means and Means+Stealing Conditions rated the agent's action as bad (Means: $M = -.55$; $CI_{95\%} = [-.90, .-.20$; Means+Stealing: $M = -.95$; $CI_{95\%} = [-1.3, -.59]$). Subjects in the Side-Effect Condition rated the action as just OK ($M = .14$; $CI_{95\%} = [-.25, .54]$).

Analysis of variance revealed a significant difference between the conditions, $F(2, 87.7)$ = 7.93, ω_p^2=.13, CI$_{95\%}$ = [.02, .26], p=.001. Planned pairwise comparisons revealed that there was a significant difference between the Means and Side-Effect Conditions (independent-sample t-test, t$_{\text{Welch}}$(89.87) = -2.56, g = -.53, CI$_{95\%}$=[-.94, -.07], p=.012, two-tailed) but no significant difference between the Means and Means+Stealing Conditions (independent-sample t-test, two-tailed, t$_{\text{Welch}}$(81.81) = 1.57, g = .34, CI$_{95\%}$=[-.09, .77], p=.12, two-tailed). See Fig. 4 for a graph of the distributions of Action Ratings (plot generated using the ggstatsplot R package, Patil, 2018).

$F_{\text{Welch}}(2,87.17) = 7.93$, $p = 0.001$, $\widehat{\omega_p^2} = 0.13$, CI$_{95\%}$ [0.02, 0.26], $n_{\text{obs}} = 134$

$p_{\text{Holm-corrected}} = 0.001$

$\widehat{\mu} = 0.14$

$\widehat{\mu} = -0.55$

$\widehat{\mu} = -0.95$

Action Rating

Means (n = 42)   Means+Stealing (n = 42)   SideEffect (n = 50)

Condition

$\log_e(\text{BF}_{01}) = -4.29$, $\widehat{R^2}_{\text{median}}^{\text{posterior}} = 0.11$, CI$_{95\%}^{\text{HDI}}$ [0.03, 0.21], $r_{\text{Cauchy}}^{\text{JZS}} = 0.71$

Pairwise test: **Games-Howell test**; Comparisons shown: **only significant**
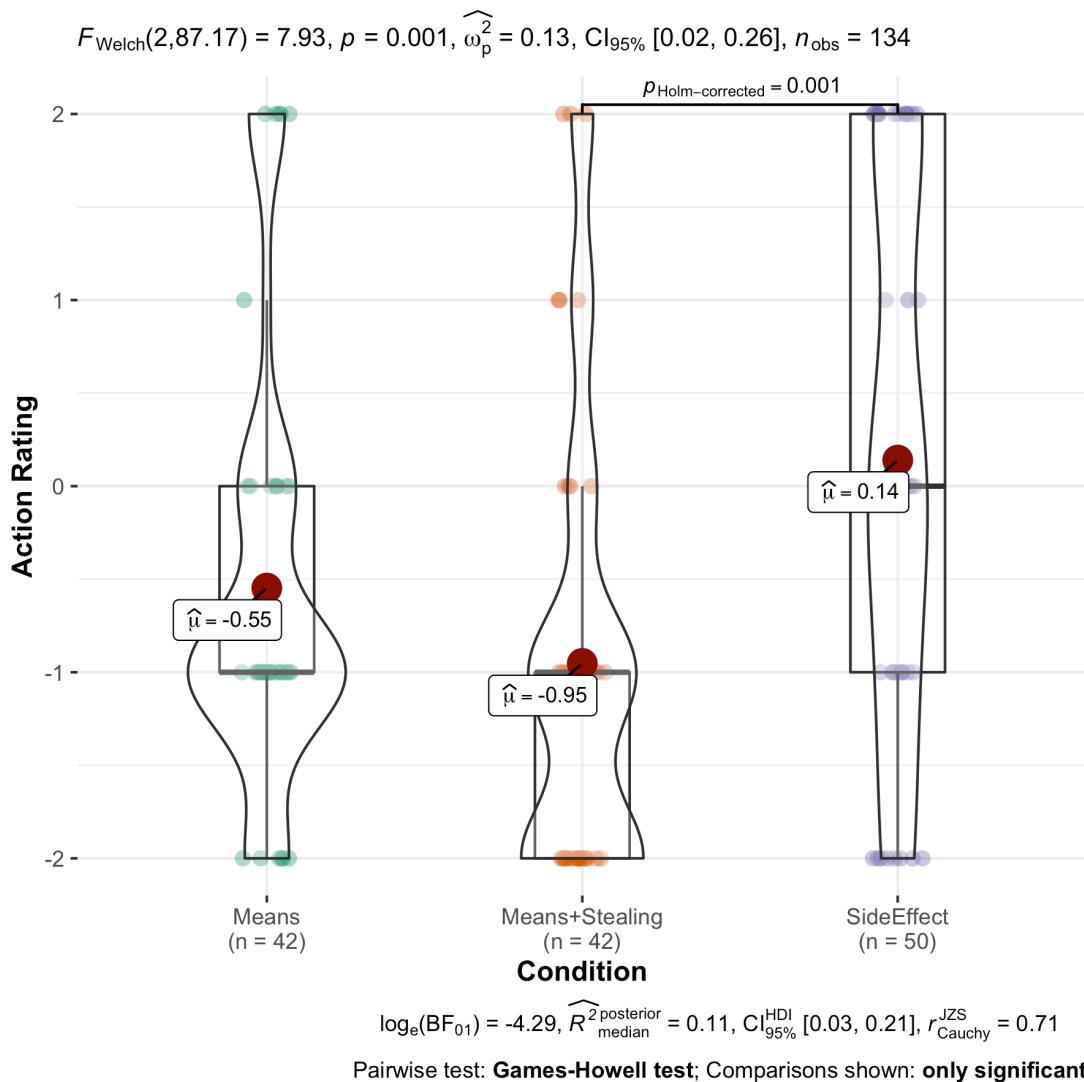
Fig. S7

Distributions for Action Ratings for the three conditions in Study 1. Subjects responded to the question, "Was what Sally did good, bad or just OK?" Likert scale ratings were scored as follows: really bad = -2; a little bad = -1; just OK = 0; a little good = 1; really good = 2. Each individual subject response is plotted and data means labeled and graphed as red dots.

Study 2: Materials and Methods

<u>NOTE</u>: Helps Others Condition for Study 2 is identical to the Means+Stealing Condition of Study 1.

Helps Victim Condition
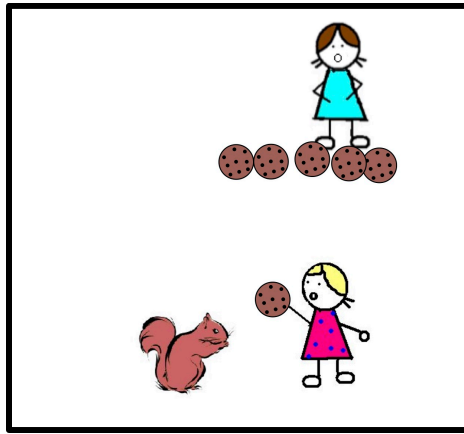
Selected scene from Helps Victim Condition of Study 2.

This is a story about Katy. And Katy is playing in the park. There is another kid in this story too. This kid is eating cookies! She has so many cookies! But ut oh, here comes a mean, sneaky squirrel. He likes to eat other people's food. That squirrel is coming over here and -- what do you think he's going to do? That's right, he wants to come over here and eat all this kids' cookies. And how is this kid going to feel if the squirrel eats her cookies? That's right, she'll feel sad. So Katy knows what the squirrel is going to do. Katy knows that the squirrel is going to go eat that kid's cookies and make her feel sad. So let's see what she does!

Katy waves to this girl. Katy waves to her and says, "Hey, come over here!" And so that girl picks up her cookies and she comes over. And Katy knows that when the girl gets there, that she is going to take her cookie and feed it to the squirrel. And the squirrel will eat her cookie. So that girl is sad because she doesn't get to eat that cookie. But then she is not sad, because she gets to eat the rest of her cookies.

Ok, let's watch that again from the beginning. [Replay the animation.]

In the beginning, do you remember where the squirrel was going to go?

How was that kid going to feel?

Control questions (subjects are excluded for incorrect answers):

And then what did Katy do?

What did the girl do?

Then what did Katy do?

What did the squirrel do?

How did that girl feel?

Then what did that girl do?

How did she feel then?

Test questions:

So that's the end of the story, but I'm wondering about something. I'm wondering about Katy and what she did. In this story, Katy took this girl's cookie and fed it to the squirrel.

[Should Question] Should she have done that?

[Action Rating] Can you show me on the pink scale? Was what Katy did, good, bad, or just OK?

Note on Test Questions: If a child said "I don't know" or something irrelevant, we prompted them with the question again. For example, "What do you think? Should she have done that?" If subjects responded with two different answers at once, (for example, "yes and no" or "good and bad"), we would ask (for example), "If you had to choose one, what would you say? Should she have done that?" If a subject still failed to provide an answer, they were excluded from the analysis.

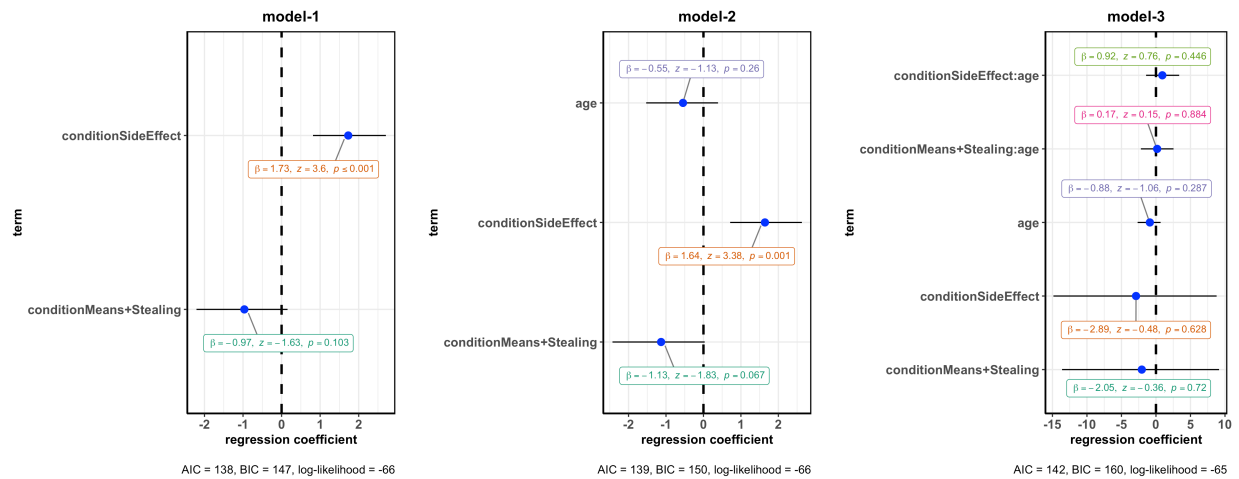# 3. Supplemental Figures for Studies 1 & 2

Fig. S8: Comparison of three possible models for Study 1 analysis of the Should Question. Model 1 includes only condition as a predictor. Model 2 includes age as well as condition. Model 3 includes the age x condition interaction. Model 1 fits the data best on AIC and BIC measures. Model 2 shows that there is no main effect of age. Model 3 shows that there is no significant interaction with age. Model comparisons here and throughout the paper and supplemental materials were conducted in R with the package ggstatsplot (Patil, 2018).
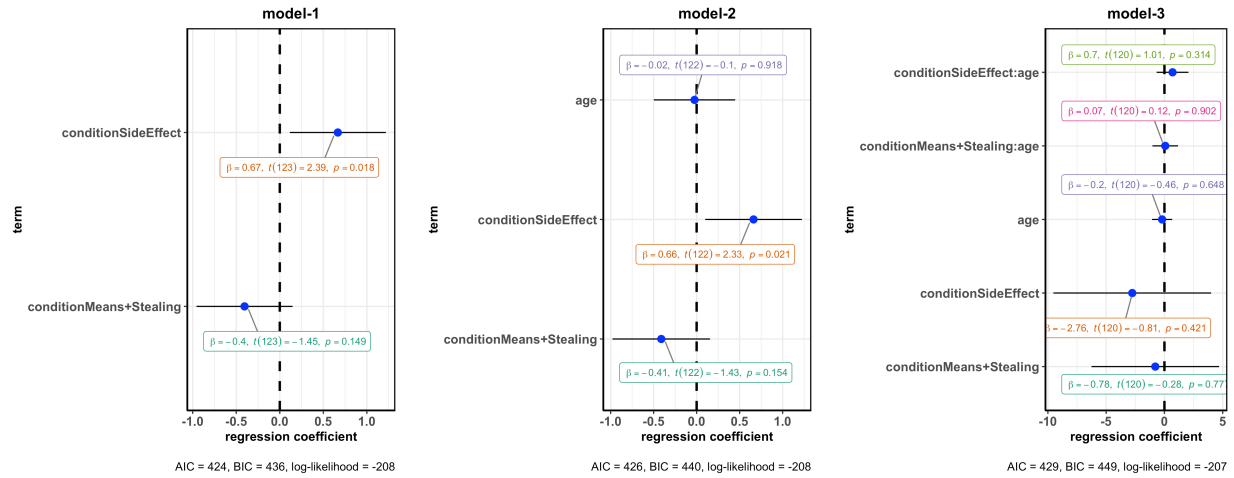
Fig. S9: Comparison of three possible models for Study 1 analysis of the Action Rating. Model 1 includes only condition as a predictor. Model 2 includes age as well as condition. Model 3 includes the age x condition interaction. Model 1 fits the data best on AIC and BIC measures. Model 2 shows that there is no main effect of age. Model 3 shows that there is no significant interaction with age.
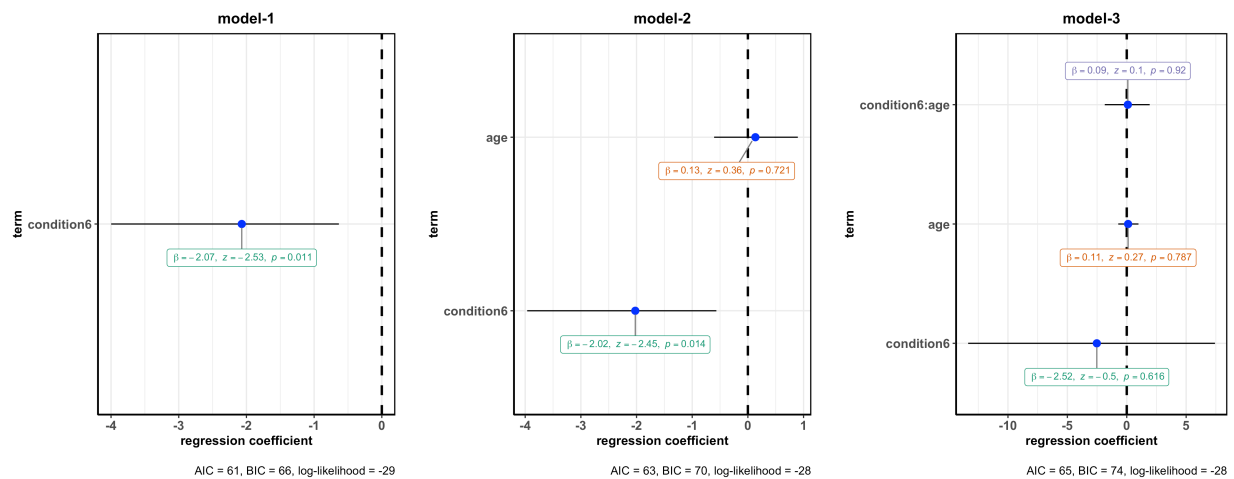

Fig. S10: Comparison of three possible models for Study 2 analysis of the Should Question. Model 1 includes only condition as a predictor. Model 2 includes age as well as condition.

Model 3 includes the age x condition interaction. Model 2 shows that there is no main effect of age. Model 3 shows that there is no significant interaction with age. Model 1 scores the best on AIC and BIC.
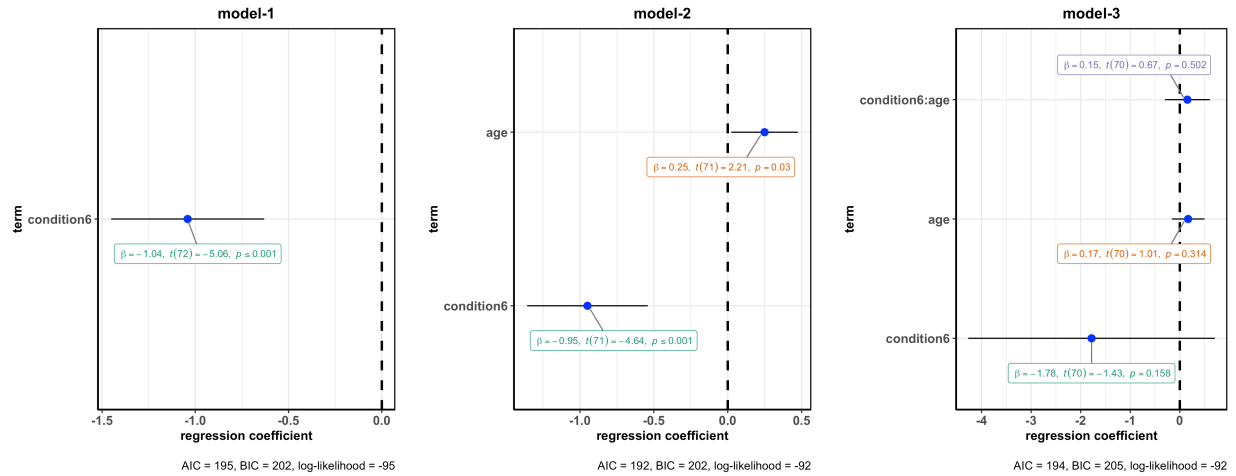

Fig. S11: Comparison of three possible models for Study 2 analysis of the Action Rating. Model 1 includes only condition as a predictor. Model 2 includes age as well as condition. Model 3 includes the age x condition interaction. Model 2 fits the data best on AIC and BIC measures. Model 2 shows that there is a significant main effect of age. Model 3 shows that there is no significant interaction of condition with age.

4. Methods, Results & Figures for Study 2b

Methods

    Procedure. Study 2b included only the Helps Victim Condition of Study 2. Responses were compared with those of Subjects in the Means+Stealing Condition of Study 1. Study 2b was conducted before Study 2 (which is reported in the main manuscript); the results of Study 2 replicate the findings of Study 2b with full randomization to both conditions.

    Subjects. Forty-two subjects between the ages of 3.5 and 6.2 years participated in the study (M=4.88 years; SD=.70; 5 three-year-olds, 20 four-year-olds, 14 five-year-olds, 2 six-year-olds), 23 of which were girls. Eighteen additional children began the study but were excluded, 7 for failing scale training/screening, 5 for failing control questions, 5 for parent intervention or distractions, 1 for experimenter error. We stopped data collection after getting data from 42 children to match the sample size in Study 1. A power analysis shows that this sample size was appropriate assuming a moderate effect size (0.4, similar to that reported in Study 1), a desired power of 0.9, and an α-level of .05. We computed that a sample of 66 subjects total (over the two conditions being analyzed) was necessary to find a difference between the two conditions with a χ^2 analysis for subjects' should judgments. The Action Rating was of secondary importance. According to our analysis, a sample of 172 subjects was necessary to find a difference in means between the two conditions with a two-tailed t-test (assuming an effect size similar to that of Study 1, d=.5, power of .9 and α-level of .05). However, as reported below, we did indeed discover the effect we were expected. (Power analysis was done with G*Power 3, see Faul, Erdfelder, Lang, & Buchner, 2007). The subjects were recruited from the Boston Children's Museum. Approximately 82% of visitors are from New England; 29% receive free or

discounted admission. Parental consent was obtained for all participants. Testing was conducted at the museum.

Results

Subjects were asked the same two test questions as in Study 1, the Should Question ("Should she have done that?") and the Action Rating ("Was what Sally did good, bad, or just OK?"). 38.1% of subjects (16 out of 42) responded "yes" to the Should Question, $CI_{95\%} = [23\%, 53\%]$. Subjects rated the agent's action as a little bad ($M = -.36$; $CI_{95\%} = [-.72, 0]$).

We compared answers to the test questions in Study 2b (where the harmful means helps the victim) to answers to the test questions in the Means+Stealing Condition of Study 1 (where the harmful means helps others).

We compared three logistic regressions to predict answers to the Should Question. Model 1 includes only study (Means+Stealing Condition of Study 1 or Study 2b) as a predictor. Model 2 includes age as well as study. Model 3 includes the age x study interaction. Given that Model 1 and Model 2 score similarly on AIC and BIC and that there were no significant effects of age in the other models (Fig. S10), the rest of the data analysis was conducted with χ^2 tests. Subjects were significantly more likely to answer "yes" to the Should Question in Study 2b as compared to the Means+Stealing Condition in Study 1, $\chi^2_{Pearson}$ (1, $n = 84$) = 7.68, BF = 14.0, $V_{Cramer} = 0.30$, $CI_{95\%} = [0.12, .54]$, $p = .006$, two-tailed. See Fig. 5.

We compared three linear regressions to predict answers to the Action Rating. Model 1 includes only study as a predictor. Model 2 includes age as well as study. Model 3 includes the age x study interaction. Given that Model 1 scored best on AIC and BIC and that there were no significant effects of age in the other models (Fig. S11), the rest of the data analysis was conducted with t-tests. Pairwise comparison revealed that there was a significant difference

between the Action Ratings in Study 2b and those in the Means+Stealing Condition of Study 1 (independent-sample t-test, $t(82) = -2.28$, BF = 2.1, $g = -.49$, $CI_{95\%} = [-.93, -.06]$, p=.025, two-tailed; see Fig. S12 for a graph of the distributions of Action Ratings).

As our preregistration document states, we deemed Study 2 a replication of Study 2b because the Bayes Factor of Study 2 was above 3 (BF = 19.7). However, Study 2 is also considered to have replicated Study 2b if p-values are considered (p=.005 for Study 2) or if effect-size is considered the appropriate measure to determine replication; the effect sizes have overlapping confidence intervals (Replication effect size: VCramer = 0.33, CI95% [0.12, .53]; Original effect size: VCramer = 0.30, CI95% = [0.12, .54]).
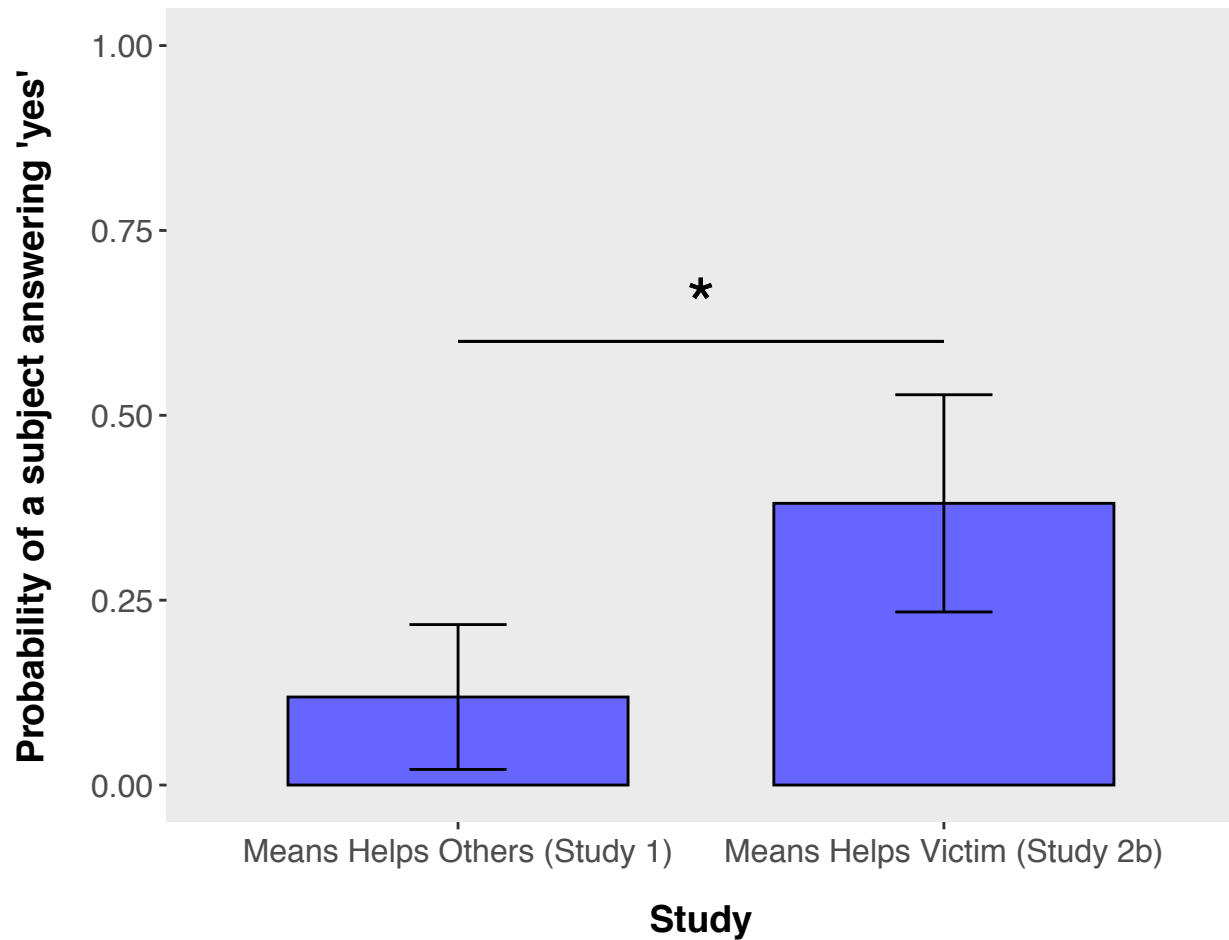
Fig. S12.

Percent of subjects responding *yes* to the question, "Should she have done that?" as a function of study. The Means+Stealing Condition in Study 1 (in which the means helps others) is compared to the Study 2b case where the means (+stealing) helps the victim. Error bars are 95% confidence intervals. The asterisks (*) indicates the significant difference between the two conditions.
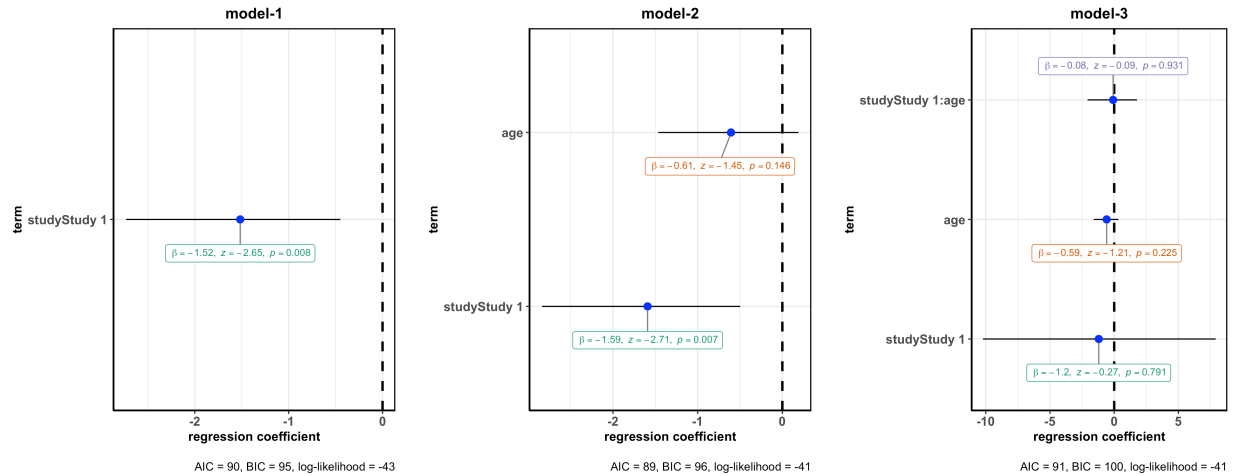
* $p < .01$

Fig. S13: Comparison of three possible models for Study 2b analysis of the Should Question.

Model 1 includes only study (Study 2b or Means+Stealing Condition of Study 1) as a predictor.

Model 2 includes age as well as study. Model 3 includes the age x study interaction. Model 2

shows that there is no main effect of age. Model 3 shows that there is no significant interaction

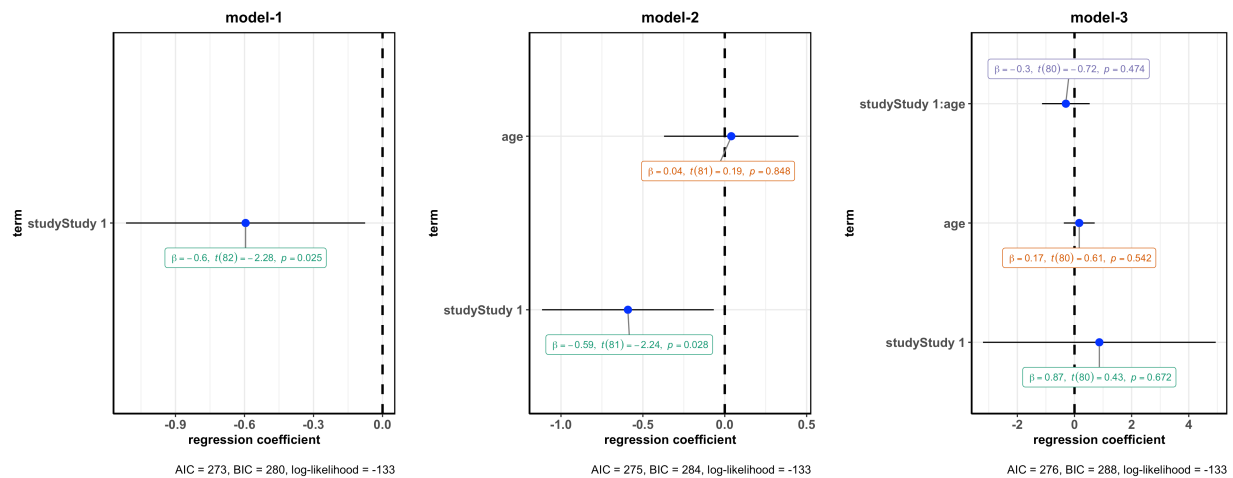with age. Model 1 and Model 2 score similarly on AIC and BIC.

Fig. S14: Comparison of three possible models for Study 2b analysis of the Action Rating.

Model 1 includes only study (Study 2b or Means+Stealing Condition of Study 1) as a predictor.

Model 2 includes age as well as study. Model 3 includes the age x study interaction. Model 1

fits the data best on AIC and BIC measures. Model 2 shows that there is no main effect of age.

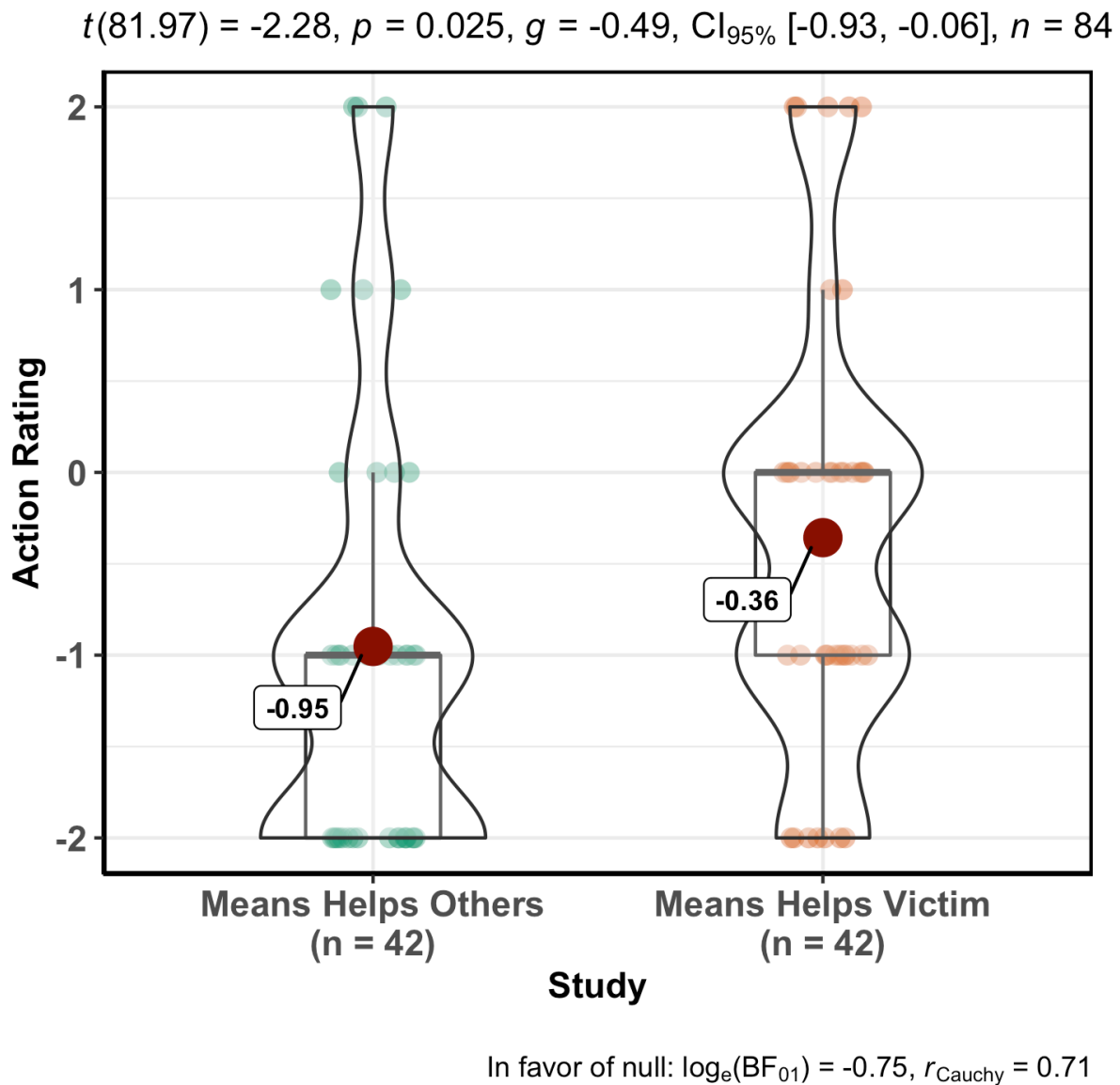Model 3 shows that there is no significant interaction with age.

$t(81.97) = -2.28, p = 0.025, g = -0.49, CI_{95\%}\ [-0.93, -0.06], n = 84$

In favor of null: $\log_e(BF_{01}) = -0.75, r_{Cauchy} = 0.71$

Fig. S15: Distributions for Action Ratings for Study 2b compared with the distribution of Action

Ratings for the Means+Stealing Condition in Study 1. Subjects responded to the question, "Was

what Sally did good, bad or just OK?" Likert scale ratings were scored as follows: really bad = -2; a little bad = -1; just OK = 0; a little good = 1; really good = 2.