



Children's cost-benefit analysis about agents who act for the greater good

Zoe Finiasz^{a,*}, Montana Shore^b, Fei Xu^c, Tamar Kushnir^a

^a Department of Psychology & Neuroscience, Duke University 417 Chapel Drive, Box 90086, Durham, NC 27708, USA

^b Department of Psychological & Brain Sciences, Boston University 64 Cummings Mall, Boston, MA 02215, USA

^c Department of Psychology, University of California, 2121 Berkeley Way, Berkeley, CA 94720, USA

ARTICLE INFO

Keywords:

Child development
Moral judgment
Greater good
Cost-benefit analysis
Personal cost
Collective consequence

ABSTRACT

Acting for the greater good often involves paying a personal cost to benefit the collective. In two studies, we investigate how children ($N = 184$, $M_{\text{age}} = 8.02$ years, $SD = 1.15$, Range = 6.00–9.99 years) use information about costs and consequences when reasoning about agents who act for the greater good. Children were told about a novel community, in which individuals could pay a cost to prevent a consequence (e.g., holding up an umbrella to prevent rain from flooding the village). In Study 1, children saw two scenarios, one where costs were minor and consequences were major, and one where the opposite was true (major cost, minor consequence). Children in the former condition expected more agents to engage in costly behavior and judged refusal to engage in costly behavior as less permissible. In Study 2 we separately manipulated cost and consequence to see which factor influences children's judgments most – cost or consequence. Here, children expected agents to pay a minor cost regardless of consequence, and only expected agents to pay a major cost when consequence was also major. In their permissibility judgments, children judged refusal to engage in costly behavior to be less permissible when consequences were major than when they were minor, regardless of cost. These findings suggest that children are making principled judgments about acting for the greater good – both cost and consequence determine when we are expected to act, but consequence seems to be a particularly key factor in deciding when inaction is permissible.

1. Introduction

Acting for the greater good can be costly. We receive painful vaccination shots to keep our community safe, spend hours in line to vote in elections, and go out of our way to recycle. Yet in all of these instances, a single individual's behavior is not enough to have an impact. One vaccination, one vote, or one recycled soda can is not enough to make a difference, so why do we bother paying these costs at all? People often evaluate one individual's behavior by asking what would happen if everyone else behaved similarly – this is known as *universalization* (Levine, Kleiman-Weiner, Schulz, Tenenbaum, & Cushman, 2020). For example, if no one gets vaccinated, we will fail to achieve herd immunity, leaving our whole community vulnerable to illness. Why then, is there so much variation in people's decision to get vaccinated? For some, vaccination may simply be too steep of a price to pay, as many hold the belief that vaccination is more harmful than the disease it is meant to prevent. On the other hand, many choose to vaccinate because they believe it is a low-cost way to prevent a dire consequence, namely a severe disease that could impact the entire community (e.g., Brewer

et al., 2007; Gidengil, Chen, Parker, Nowak, & Matthews, 2019; Karlsson et al., 2021; Wise, Zbozinek, Michelini, Hagan, & Mobbs, 2020). This highlights a key feature of our decisions about when to act for the greater good: the tradeoff between cost and consequence.

Many of the norms we teach young children follow the same logic of universalization. For example, children learn to raise their hands to talk in classrooms, or stand in line and wait their turn, because multiple children talking at once or all trying to be first would cause harm and create chaos for the group. Do children evaluate individuals in these scenarios in light of the hypothetical consequences to the collective? To answer this, a recent study by Levine et al. (2020) examined how 4- to 11-year-old children and adults make moral judgments about an individual who chooses not to engage in costly behavior to act on behalf of the greater good. In one vignette, children and adults heard a story about a fictional character who picked up a stone from a path. Participants were told that if too many stones were removed the path would disappear. Across two conditions, participants then learned that the character who took the stone was either the only one who wanted a stone (*low interest condition*), or they learned that many others also wanted stones

* Corresponding author.

E-mail addresses: zoe.finiasz@duke.edu (Z. Finiasz), mjshore@bu.edu (M. Shore), fei_xu@berkeley.edu (F. Xu), tamar.kushnir@duke.edu (T. Kushnir).

(*high interest condition*). Even though the act on its own was harmless and the consequences to the collective good were hypothetical, adults and 4- to 11-year-old children judged the action in the high interest condition to be less permissible. These findings demonstrate how the logic of universalization guides children's moral reasoning. Levine et al. (2020) found that children, like adults, not only look beyond the immediate consequences of actions, but also potential unrealized consequences for the collective when judging the permissibility of an action.

We suggest that to fully elaborate this view, whether it is a viable account of early-emerging moral judgment and whether it is a useful account of norm creation, we must resolve how tradeoffs are considered. This is because even if we agree that any individual action can be judged on the grounds that it may have collective consequences, we can productively *disagree* about either the personal cost of the action and/or the severity of the consequences to the group. As the vaccination example above suggests, these disagreements could lead us to excuse actions with collective consequences if we feel the personal cost is too high, or the consequences are too low, or some combination thereof. Thus, the principle of universalization does not alone explain *why* actions are judged to be more or less permissible, nor *why* some of us view the same actions as candidates for moralization while others do not.

To begin to address this question, we ask whether variations in personal cost and collective consequences inform children's tendency to universalize – that is, we ask how children judge single actions that are only harmful in aggregate based on both the cost of the action and the degree of collective harm. We explore this question across three dimensions. First, we ask how cost and consequence shape children's predictions about *how many others* will act. Second, we ask how tradeoffs between costs and consequences change children's evaluative judgments of individual actions, both when consequences to the collective are as of yet unrealized (as in Levine et al., 2020), and again when consequences have been realized. Finally, after enough individuals have acted to pass the threshold for negative consequences, we compare children's evaluations of individuals to evaluations of the whole group. We include this as a way of probing whether children equate individual and collective responsibility.

There is a growing literature showing the early-emerging influence of cost on children's social inferences; from infancy through childhood, children view costly actions as an indicator of other's subjective values, goals, and preferences (e.g., Bridgers, Jara-Ettinger, & Gweon, 2020; Jara-Ettinger, Floyd, Tenenbaum, & Schulz, 2017; Jara-Ettinger, Gweon, Schulz, & Tenenbaum, 2016). When the physical cost of an action is high (i.e., having to climb over a barrier to retrieve an object), infants infer that the agent values the goal more (Liu, Ullman, Tenenbaum, & Spelke, 2017). Two-year-old children inferred an agent's desire to behave prosocially was higher when it was physically costly, compared to when the same act came at no cost (Jara-Ettinger et al., 2016; Jara-Ettinger, Tenenbaum, & Schulz, 2015). When 5- to 6-year-old children were asked to predict another person's preferences, they incorporated information about how costly an item was to obtain: the more someone went out of their way to obtain a certain item, the more children believed that this cost was paid to achieve a subjectively better reward (Jara-Ettinger, Gweon, Tenenbaum, & Schulz, 2015).

Increasingly from early to middle childhood, information about costs also shapes children's social and moral evaluations. For instance, in one prior study, when two people refused to engage in costly behavior to help another person, toddlers indicated that the person who would have paid a higher cost to help was nicer (Jara-Ettinger, Tenenbaum, & Schulz, 2015). Another study showed that by 5, children used information about costs to make judgments about who is more competent (Jara-Ettinger, Gweon, et al., 2015). Further, by 6- to 7-years, children recognized an actor as more praiseworthy when it was physically costly for him to fulfill a social obligation compared to when that same action came at no cost (Zhao & Kushnir, 2022).

Much like their understanding of costs, children's attention to actions with negative moral consequences begins in infancy (Hamlin,

Wynn, & Bloom, 2007). Toddlers recognize that it is wrong to cause a negative interpersonal consequence and will negatively judge an actor who has caused harm to another person (Hamlin, Wynn, Bloom, & Mahajan, 2011). Preschool children are more upset if someone causes negative consequences intentionally than accidentally (Josephs, Kushnir, Gräfenhain, & Rakoczy, 2016), but still tend to evaluate transgressions negatively, regardless of the actors' intentions (Josephs et al., 2016; Killen, Mulvey, Richardson, Jampol, & Woodward, 2011). Relevant to our current investigation, one recent study has found that children use information about consequences to guide their reasoning about cost-benefit tradeoffs. For example, when given information about someone engaging in costly giving, 6- to 7-year-old children's generosity ratings were sensitive to both the cost to the giver and the recipient's need (Radovanovic et al., 2023). Similar work with toddlers has shown that 16-month-olds show a preference for agents who help others: when presented with two agents, toddlers showed a preference for the one whose actions helped someone in greater need (Woo, Liu, Gweon, & Spelke, 2024). Together these studies show that when children are making evaluations about a helper, they are considering the impact that the helpers' action has. This work suggests that at least by age 6, children consider tradeoffs between costs and consequences in interpersonal contexts.

In addition to Levine et al. (2020) a small number of prior studies have explored children's understanding of *collective* consequences. For example, Smith and Warneken (2016) found that 4-year-old children reason that sometimes the actions of one individual can lead to consequences for everyone, but it's not until age 6 that they separate collective consequences from collective punishment. Along similar lines, Thomas, Kelsey, and Vaish (2024) showed that 5- to 7-year-olds negatively evaluated those whose individual actions bring about collective punishments for a group, citing unfairness. Finally, there is some evidence to suggest that between 5- and 7-years, children also reason about collective *responsibility*; previous work has shown that children understand that the responsibility of one person's actions may be shared collectively amongst their group members (Over, Vaish, & Tomasello, 2016), even going so far as to apologize for the actions of another ingroup member (Bennett & Sani, 2008) or lying to protect in-group transgressors (Misch, Over, & Carpenter, 2016).

In sum, the research reviewed above shows that toddlers' and preschoolers' social inferences are based on considerations of costs and consequences; it is not until age 6 that children consider trade-offs between cost and consequence in their social evaluations. Moreover, all such prior work has only presented children with cases where an individual's action has direct consequences for another individual (or for the collective). In cases where one must act for the greater good, our predictions and judgments must encompass more than merely personal cost versus collective consequence. An essential component of these threshold problems is that when individuals in these scenarios are deciding what costs are worth paying, their decision to engage in costly behavior is informed both by their cost-benefit analyses *and also* what they expect others in their group to do. Thus, our questions – whether personal cost and collective consequence influence children's tendency to universalize, and relatedly whether cost and consequences influence moral judgments when a single action has *no* negative consequences of its own, remain unanswered.

Across two studies, we explore these questions in 6- to 9-year-old children. We use novel vignettes (similar to Levine et al., 2020) where children were told about a collective consequence that only occurs if a certain number of individuals act a certain way. Rather than vary *interest* (high or low) in the action, we vary the personal *cost* of the action to the individual and the *consequences* of the action to the group (minor or major). We first asked children to predict interest (how many people will act) based on cost and consequence. Most importantly, to see if the tendency to universalize is influenced by cost and consequence, we asked them to make judgments about a single action prior to any consequences occurring (as in Levine et al., 2020). To further probe their

moral judgments, we asked them to evaluate both a single actor and the whole group of agents who together acted to cause collective harm (e.g., passed the threshold for collective consequence). In Study 1, we began by contrasting the extremes – scenarios where personal costs are minor and collective consequences major, versus scenarios where costs are major and consequences minor. We also include two types of vignettes; one adapted from Levine et al. (2020) where the agents must engage in costly behavior by resisting a desire (psychological cost) and a novel one created for our study where agents must engage in costly behavior by lifting a heavy object (physical cost). We follow this by exploring the intermediate cases in Study 2, where we separately manipulated cost and consequence to see how each factor influences children's tendency to universalize and their predictions and moral judgments.

2. Study 1

The present study investigates whether children engage in cost-benefit analyses when making predictions and judgments about agents who must engage in costly behavior to prevent a collective consequence. Here we compare children's responses in cases where the contrast between cost and consequence – Minor Cost/Major Consequence, and Major Cost/Minor Consequence – is highest. We began with these high-contrast scenarios to investigate whether or not children engage in cost-benefit analysis when reasoning about the greater good, with the goal of disambiguating whether children prioritize cost, consequence, or both in Study 2. As such, Study 1 explores how children use both personal cost and collective consequence together in their reasoning. For each case, children were asked to *predict* the number of agents who would engage in costly behavior and make *permissibility judgments* about agents' refusal to engage in costly behavior, both prior to, and following consequences. If children's tendency to universalize is sensitive to considerations of both personal cost and collective consequence together, we expect that children will rate refusal to engage in costly behavior as more permissible when the cost of doing an action is major but the collective consequence it prevents is minor. Additionally, we look at whether personal cost and collective consequence together influence predictions about *how many* agents will engage in costly behavior (a measure of interest), and moral evaluations of both individuals and groups (measures of individual responsibility vs collective responsibility).

2.1. Method

2.1.1. Participants

Sixty-four 6- to 9-year-old children (27 girls, 37 boys) in a medium-sized city in the Southeastern US ($M_{age} = 7.96$ years, $SD = 1.18$ years, Range = 6.03–9.94 years) participated in the study. Children were recruited from a developmental research participant database, a local science museum, and through social media (i.e., Facebook, Instagram, and Twitter). An additional 5 participants were tested but excluded from our analyses (1 failed memory checks and 4 did not complete the study).

Of the respondents who reported on their child's ethnicity (95.31 %, $N = 61$), 70.49 % identified as White ($n = 43$), 8.20 % as Hispanic or Latino ($n = 5$), 8.20 % as multiracial ($n = 5$), 6.56 % as Asian ($n = 4$), and 6.56 % as Black or African American ($n = 4$). Of the caregivers who reported their income (92.19 %, $N = 59$) and education (93.75 %, $N = 60$), most had an average household income of over \$100,000 USD (71.19 %, $n = 42$) and had a graduate/professional school degree (66.67 %, $n = 40$).

2.1.2. Vignettes

Using a 2×2 mixed design, children were randomly assigned to a Cost condition (Psychological, Physical; between-subjects). In the *Psychological Cost* condition, children heard about agents who had to pay a psychological cost by resisting a desire to take stones from a path that led to a park (adapted from Levine et al., 2020). We also designed a similar vignette for the *Physical Cost* condition: children heard about

agents who had to pay a physical cost by holding up an umbrella to protect their village from rain. In both vignettes, children learn about a fictional scenario featuring fictional creatures. We elected to use these fictional vignettes because our goal was to explore how children make predictions and judgments about the greater good without the influence of any pre-existing biases or personal experiences where children themselves may have made a decision about acting for the greater good. Because children were asked to predict and judge others' behaviors, a fictional scenario allowed us to avoid any situation where children may have an existing expectation about how one ought to act.

In both conditions, children heard two vignettes that manipulated the degree of personal cost and collective consequence (Major Cost/Minor Consequence, Minor Cost/Major Consequence; within-subjects). Together, children heard two stories, one where the personal cost was major (Psychological: strong desire to pick up the stones; Physical: very heavy umbrellas) and the collective consequence was minor (Psychological: path was not necessary to find a park; Physical: some puddles in the village), and a second vignette where the personal cost was minor (Psychological: did not care about the stones; Physical: umbrellas were easy to carry) and the collective consequence was major (Path: needed path to find a park; Umbrella: village would get flooded and destroyed) (See Fig. 1). To illustrate the threshold for collective consequence in each vignette, children learned that if one agent alone refused to engage in costly behavior, no consequence would occur, but if "too many" agents refused to engage in costly behavior, everyone would experience the consequence. All vignettes showed that a majority of characters (6 out of 10) had to refuse to engage in costly behavior for any collective consequence to occur, however this number was never explicitly stated.

2.1.3. Test questions

After each vignette, children answered four test questions: one Prediction question and three Permissibility Judgment questions. The prediction question asked children to estimate how many agents (out of 10) would pay the specified cost. The Permissibility Judgment questions asked children to judge an agent (or agents) who refused to engage in costly behavior (Likert scale: Really not okay – Really okay). In the *Lone Agent* question, children judged a lone agent who refused to engage in costly behavior for the greater good. In the *Many Actors* questions, children learned about a group of agents who refused to engage in costly behavior – here, a majority (6 agents – though again this number was never explicitly stated) refused to engage in costly behavior, and the collective consequence was realized. Children were asked to make two judgments in this case – one about a single actor from the group of actors (*Single of Many*), and one about the group of six agents who refused to engage in the costly behavior (*Whole Group*).

2.1.4. Procedure

Children were interviewed via Zoom or in-person. Prior to the interview, children received training to familiarize them with the scales used in both kinds of test questions with a set of unrelated examples. After training questions were completed, children heard one vignette, answered two memory questions. One memory question asked children to recall the cost that the characters had to pay, and one asked them to recall the consequence that occurred when enough of the characters refused to engage in costly behavior. Children who did not answer these questions correctly still completed the test questions, but their responses were excluded from analysis. After the two memory questions, children answered the four test questions. The memory questions and test questions were repeated a second time following the next vignette. The order of the vignettes was counterbalanced, and the order of the memory & test questions was fixed.

2.1.5. Coding

For the Prediction Questions, children's responses (0–10) were recorded. Permissibility Judgments were given on a 0–3 Likert scale with lower scores representing less permissibility (0: Really not okay, 1:

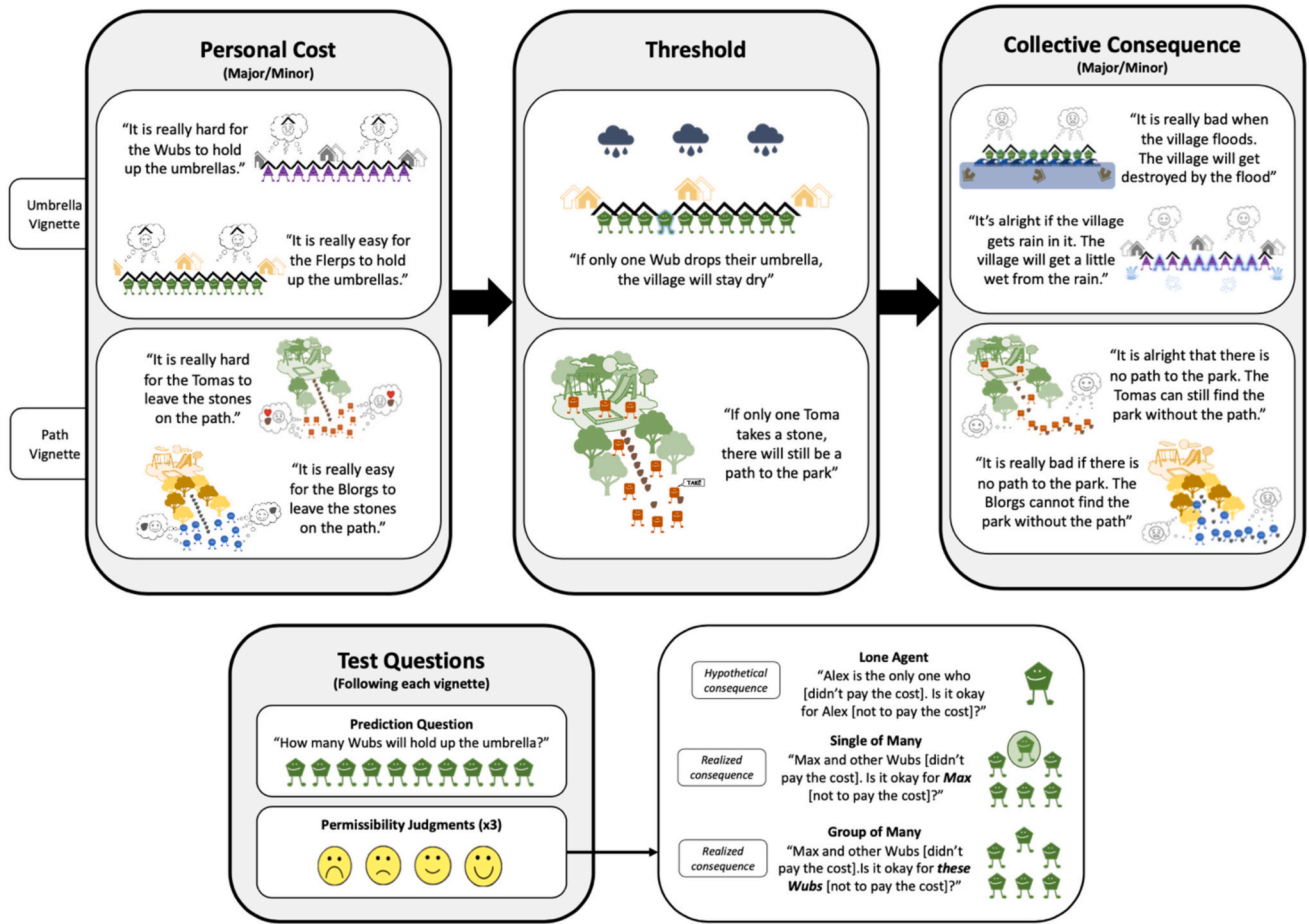


Fig. 1. Vignettes and test questions for Studies 1 and 2. The Path Vignette is only included in Study 1; the Umbrella Vignette is used in both Studies 1 and 2.

A little not okay) and higher scores indicating more permissibility (2: A little okay, 3: Really okay). Each child answered two prediction questions (one for each Cost-Consequence Condition) and 6 Permissibility Judgments (three for each condition).

2.2. Results

2.2.1. Prediction questions

To investigate children’s predictions about how many agents would engage in costly behavior to prevent a collective consequence, we ran a linear mixed effects model with Vignette (Psychological, Physical; between-subjects), Cost-Consequence (Minor Cost/Major Consequence, Major Cost/Minor Consequence; within-subjects), and Trial Order as predictors, and Age (in months) as a covariate and ID as a random effect. To perform these and all following analyses, we began with a full model including the relevant main effects and interactions and selected the simplest model to best describe our data. Information about our model selection plan and subsequent model comparisons can be found on our OSF page.

We began by comparing this model to a null model (with only Age, Trial Order, and Participant ID as predictors). Results from this comparison indicated that Cost-Consequence and Vignette together explained a nonzero proportion of variation in children’s predictions, $\chi^2(7, N = 64) = 48.34, p < .001$. We found a main effect of Cost-Consequence ($\chi^2(1, N = 64) = 19.61, p < .001$): children expected fewer agents to pay a major personal cost to prevent a minor collective consequence ($M_{MajMin} = 4.55/10$ agents, $SD = 3.12$) than in the opposite case ($M_{MinMaj} = 6.78/10$ agents, $SD = 3.19; t(63) = 4.51, p < .0001$; See

Fig. 2). We also found a main effect of Vignette condition ($\chi^2(1, N = 64) = 32.83, p < .001$): regardless of the degree of trade-off between cost and consequence, children predicted fewer agents would pay the psychological cost in the Psychological Condition ($M_{Psych} = 4.08/10$ agents, $SD = 2.57$) than the physical cost in the Physical condition ($M_{Physical} = 7.06/10, SD = 3.32; t(60) = 6.01, p < .0001$). There were no other significant predictors (all $ps > 0.10$). Together, these findings support the idea that children’s predictions are affected by their considerations of costs and consequences.

2.2.2. Universalization

o explore whether children universalized in their permissibility judgments, we first analyzed their responses when only one agent had refused to engage in costly behavior, and no consequences had occurred (*Lone Agent* question). We ran a linear mixed effects model with Vignette (Psychological vs. Physical) and Cost-Consequence (High Cost/Low Consequence, Low Cost/High Consequence), with Age (in months) as a covariate and ID as a random effect.

We began by comparing this model to a null model (with only Age, Trial Order, and Participant ID as predictors). Results from this comparison indicated that Cost-Consequence and Vignette together explained a nonzero proportion of variation in children’s permissibility judgments about a lone agent’s refusal to engage in costly behavior, $\chi^2(2, N = 64) = 19.60, p < .001$. There was a main effect of Cost-Consequence ($\chi^2(1, N = 64) = 19.57, p < .001$) but no other significant predictors (all $ps > 0.64$). Post-hoc analysis revealed children judged refusal to engage in costly behavior as less permissible when the cost was minor, and the potential consequences were major ($M_{MinMaj} =$

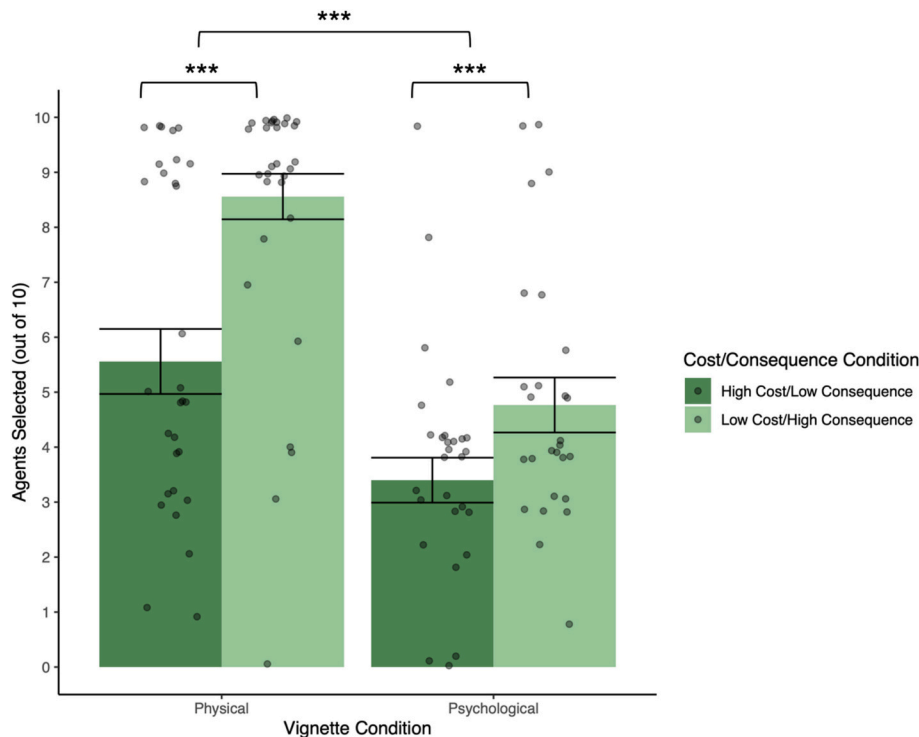


Fig. 2. Children's responses to Prediction Questions by Vignette & Cost-Consequence Conditions, Study 1. Error bars represent 1 SE. *** $p < .001$.

1.84, $SD = 0.78$) compared to the opposite case ($M_{MajMin} = 2.33$, $SD = 0.69$; $t(121) = 4.74$, $p < .001$). These findings provide evidence that children are universalizing in their permissibility judgments. Although they judged both actions as permissible, it was significantly more permissible to act when the personal costs were minor, and the hypothetical consequences were major.

2.2.3. Permissibility judgments

Next, we compared children's permissibility judgments across all three questions – including the *Lone Agent* question above (where no consequences are realized) and the two questions after consequences were realized (*Single of Many*, *Whole Group*). To do so, we ran another linear mixed effects model, this time including children's permissibility judgments about a group of agents who refused to engage in costly behavior. We used Vignette (Psychological vs. Physical), Cost-Consequence (High Cost/Low Consequence, Low Cost/High Consequence), Question Type (Lone, Single, Group; within-subjects), and Trial Order as predictors, with Age (in months) as a covariate and ID as a random effect.

We began by comparing this model to a null model (with only Age, Trial Order, and Participant ID as predictors). Results from this comparison indicated that Cost-Consequence, Vignette, and Question Type explained a nonzero proportion of variation in children's predictions, $\chi^2(4, N = 64) = 174.17$, $p < .001$. We found main effects of Cost-Consequence ($\chi^2(1, N = 64) = 77.12$, $p < .001$) and Question Type ($\chi^2(2, N = 164) = 119.09$, $p < .001$). All other predictors were not significant (all $ps > 0.27$). Post-hoc comparisons revealed that children judged refusal to engage in costly behavior as less permissible when agents were paying a minor cost to prevent a major collective consequence ($M_{MinMaj} = 0.98$, $SD = 1.00$) than in the opposite case ($M_{MajMin} = 1.78$, $SD = 1.06$; $t(317) = 9.52$, $p < .0001$). Further, when there was only one character who refused to engage in costly behavior, and a collective consequence had not been realized, children judged the refusal to engage in costly behavior as more permissible ($M_{LoneAgent} = 2.09$, $SD = 0.77$) than when they were a part of a large enough group of actors and the collective consequence had been realized ($M_{SingleOfMany} = 1.16$, $SD =$

1.07; $t(317) = 8.54$, $p < .0001$). They also judged a single actor's refusal to engage in costly behavior as more permissible than a group of agents ($M_{WholeGroup} = 0.89$, $SD = 1.07$; $t(317) = 11.38$, $p < .0001$; See Fig. 3). Children did not differ in their judgments of a single actor from the group versus the whole ($t(372) = -2.53$, $p > .01$).

Finally, we compared children's responses to the median possible ranking (1.5). For the *Lone Agent* question, children's responses in both conditions were significantly higher than the median. This suggests that children judged refusal to engage in costly behavior in each scenario as somewhat permissible [($M_{MinMaj} = 1.84$, $SD = 0.78$; one-sample- $t(63) = 3.52$, $p < .001$), ($M_{MajMin} = 2.33$, $SD = 0.69$; one-sample- $t(63) = 9.58$, $p < .001$)]. In both the *Single of Group* and *Whole Group* questions, when costs were minor and consequences were major, children rated refusal as impermissible whether they were judging an single individual or the entire group [($M_{MinMaj, SingleOfGroup} = 0.67$, $SD = 0.84$; one-sample- $t(63) = -7.92$, $p < .001$), ($M_{MinMaj, WholeGroup} = 0.42$, $SD = 0.73$; one-sample- $t(63) = -11.81$, $p < .001$)]. In the opposite case, children's responses were no different from the median score [($M_{MajMin, SingleOfGroup} = 1.64$, $SD = 1.06$; one-sample- $t(63) = 1.06$, ns), ($M_{MajMin, WholeGroup} = 1.36$, $SD = 1.15$; one-sample- $t(63) = -0.98$, ns)]. Together, these findings show that children's permissibility judgments are shaped by their cost-benefit analysis. Further, these findings show that children judge hypothetical consequences as more permissible than realized ones. Finally, these results indicate that children hold members in a group as collectively responsible for the consequences of refusing to engage in costly behavior.

2.3. Discussion

Study 1 shows that 6- to 9-year-old children jointly consider the personal cost and collective consequence in their predictions and judgments of agents who act for the greater good. When agents must pay a minor personal cost to prevent a major collective consequence, children expect more agents to engage in costly behavior (effectively expecting low interest in acting selfishly) than the opposite case (where they effectively expect high interest in acting selfishly). Importantly, children

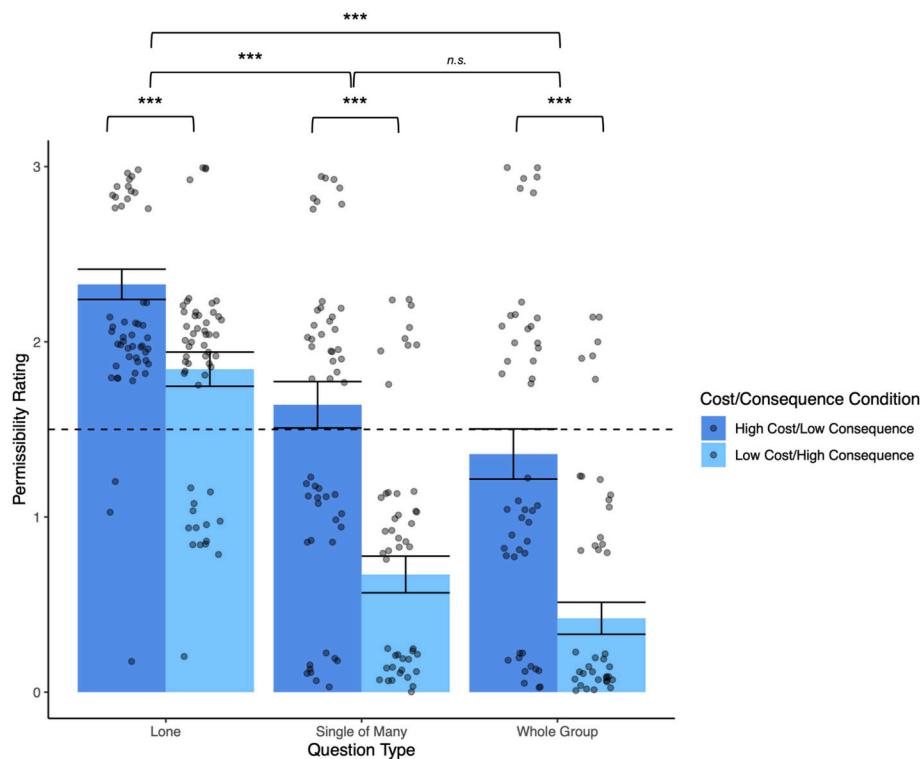


Fig. 3. Children's responses to Permissibility Judgments by Cost-Consequence Condition, Study 1. Dashed line represents the mean possible permissibility rating, error bars represent 1 SE. *** $p < .001$.

universalized more, judging single actions without any consequences as less permissible when costs to the individual were low and consequences to the collective were high. Thus, our results suggest that when children think about hypothetical collective consequences, they are more likely to evaluate single harmless actions negatively when they come at little personal cost but have the opportunity to prevent major collective harms.

Of course, once the harmful consequences are realized, children's evaluations become more negative still. But, even in the case when the consequences have already occurred, children consider personal cost and collective consequence together in their evaluations. Additionally, they evaluate each individual action as equally impermissible to the collective actions of the group. This suggests that children hold individuals responsible for group behavior, aligning with findings from prior work (Smith & Warneken, 2016). Further, this provides evidence that children's judgments are affected in some part by either personal costs, collective consequences, or both factors together.

Study 1 also investigated children's consideration of the kind of personal cost – physical or psychological. Children in the psychological cost condition showed the same pattern of results – they predicted fewer agents would pay a high personal cost to prevent a minor collective consequence – but overall differences between conditions were lower. Consistent with previous findings on this question, our results suggest that children's application of psychological costs to social evaluations may develop later than their understanding of physical costs (e.g., Starmans & Bloom, 2016; Zhao & Kushnir, 2022). An alternate interpretation of these findings can be explored across the action/inaction distinction. Children judge harm caused by inaction as more permissible than harm caused by action (Hayashi, 2015; Powell, Derbyshire, & Guttentag, 2012) suggesting that there may be different norms guiding children's inferences about how others will (or will not) act. While the kind of cost did not change children's permissibility judgments in our study, it is possible that this distinction similarly shaped children's predictions of who would engage in costly behavior. Future work can investigate this further in the context of acting for the greater good.

3. Study 2

Study 1 investigated children's cost-benefit analyses in the extreme cases – when cost is minor and consequence is major versus its opposite. From these findings, we can conclude that children consider *either* personal cost or collective consequence in their judgments and predictions but leaves an open question about which factor they prioritize in their reasoning. In Study 2, we manipulate the degree of personal cost and collective consequence separately. We use the same test questions in a mixed design, where personal cost is varied between participants and collective consequence is varied within. In Study 2, we focus on cases where the personal cost incurred is physical as these showed the clearest condition differences. Study 2 includes the same set of test questions as in Study 1.

3.1. Method

3.1.1. Participants

One-hundred-and-twenty 6- to 9-year-old children (59 girls, 60 boys, and 1 non-binary) in Durham, NC ($M_{age} = 8.04$ years, $SD = 1.14$ years, Range = 6.00–9.99 years) participated in the present study. Children were recruited from a developmental research participant database, a local museum, and through social media (i.e., Facebook, Instagram, and Twitter). An additional 11 participants were tested but excluded from our analyses (4 failed memory checks, 1 parent interference, 5 did not complete study, and 1 was outside of the age range). A preregistered a priori analysis (ANCOVA, sufficient power for main effects and interactions) determined a minimum sample size of 118 participants to obtain 80 % power. An additional two participants were added in order to achieve evenly distributed counterbalancing across conditions.

Of the respondents who reported on their child's ethnicity (97.5 %, $N = 117$), 72.65 % identified as White ($n = 85$), 11.97 % identified as multiracial ($n = 14$), 6.84 % identified as Asian ($n = 8$), 5.13 % identified as Black or African American ($n = 6$), and 3.42 % identified as Hispanic or Latino ($n = 4$). Of the caregivers who reported their income (95.83, N

= 115), most had an average household income of over \$100,000 USD (63.48 %, $n = 73$). Most caregivers had a graduate/professional school degree (58.33 %, $n = 70$).

3.1.2. Procedure

Using a 2×2 mixed design, children were randomly assigned to a Cost condition (Minor Cost, Major Cost; between-subjects). Within each Cost condition, children heard two vignettes that manipulated collective consequence (Minor Consequence, Major Consequence; within-subjects). All participants heard the same Physical Cost vignette from Study 1 (holding umbrellas to protect their village from rain). Children were interviewed via Zoom and in-person. The procedure, familiarization questions, test questions, and coding scheme were the same as in Study 1. Finally, an additional survey was distributed to parents of the participants in Study 2 asking about their own engagement in costly behavior for the greater good. Due to low response rates, the data from this survey was dropped from our analysis plan but can be found at the following link: https://osf.io/whzfm/?view_only=bf364806e63445519feab40cb505b028.

3.2. Results

3.2.1. Prediction questions

To investigate children's predictions about how many agents would engage in costly behavior to prevent a collective consequence, we ran a linear mixed effects model with Cost (Minor, Major; between-subjects), Consequence (Minor, Major; within-subjects), and their two-way interaction as predictors, with Age (in months) as a covariate and ID as a random effect.

We began by comparing this model to a null model (with only Age, Trial Order, and Participant ID as predictors) Results from this comparison indicated that Cost, Consequence, and their interaction together explained a nonzero proportion of variation in children's predictions, $\chi^2(2, N = 120) = 69.62, p < .001$. We found a significant interaction between Cost and Consequence ($\chi^2(1, N = 120) = 8.99, p = .002$; See Fig. 4) where, when personal cost was major, children expected fewer agents to act to prevent a minor consequence ($M_{MajCost,MinCons} = 5.10, SD = 2.83$) than a major one ($M_{MajCost,MajCons} = 6.88, SD = 3.11; t(118) =$

5.15, $p < .001$). However, when the personal cost was minor, children's predictions did not differ by consequences, they generally expected most agents to act ($M_{MinCost,MinCons} = 8.75, SD = 2.07; M_{MinCost,MajCons} = 9.05, SD = 2.06; t(118) = 0.87, p > .01$). There were no effects of age nor trial order (all $ps > 0.72$). Together, these findings demonstrate that children are using information about both personal costs and collective consequences to predict how many agents will act for the greater good.

3.2.2. Universalization

To explore whether children universalized in their permissibility judgments, we first analyzed their responses when only one agent had refused to engage in costly behavior, and no consequences had occurred. We ran a linear mixed effects model with Cost (Major vs. Minor) and Consequence (Major vs. Minor) with age (in months) as a covariate and ID as a random effect.

We began by comparing this model to a null model (with only Age, Trial Order, and Participant ID as predictors) Results from this comparison indicated that Cost and Consequence, together explained a nonzero proportion of variation in children's predictions, $\chi^2(2, N = 120) = 18.16, p < .001$. We found a main effect of Consequence ($\chi^2(1, N = 120) = 16.88, p < .001$) and no other significant predictors (all $ps > 0.26$). Post-hoc analyses revealed that children judged refusal to engage in costly behavior as more permissible when the potential consequences were minor ($M_{Minor} = 2.31, SD = 0.65$) compared to when they were major ($M_{Major} = 2.06, SD = 0.74, t(119) = 4.24, p < .001$). This provides evidence that children are universalizing in their permissibility judgments about a lone agent's refusal to engage in costly behavior. Although the realized outcome was the same in both scenarios, it was less permissible to refuse to engage in costly behavior when the hypothetical collective consequences were major. Cost did not influence children's judgments of a lone agent.

3.3. Permissibility judgments

Next, we explored whether children's permissibility judgments differed after enough agents refused to engage in costly behavior, and a collective consequence was realized. To do so, we ran another linear mixed effects model, this time including children's permissibility

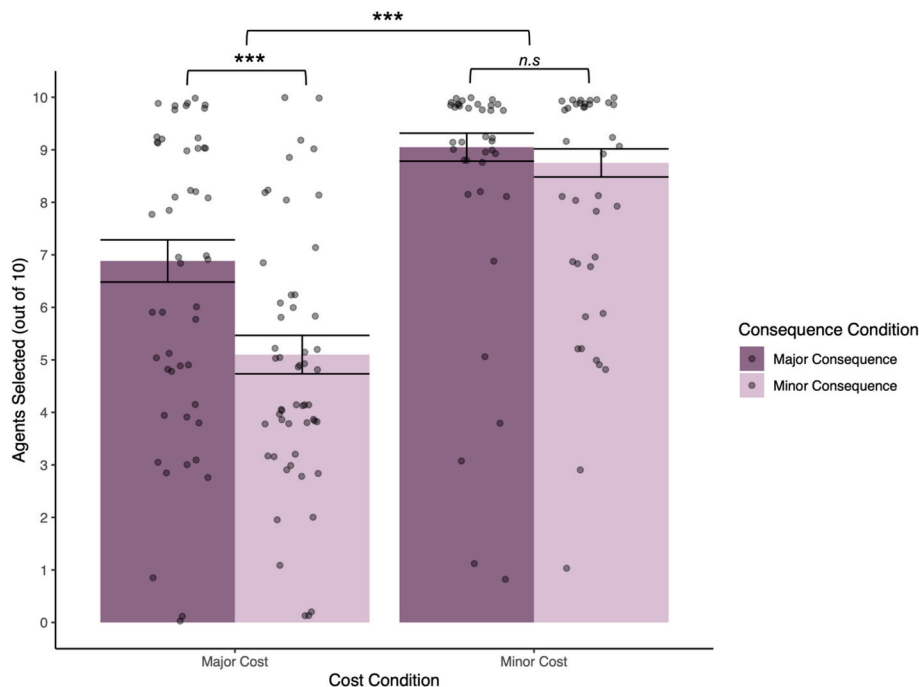


Fig. 4. Children's responses to Prediction Questions by Cost & Consequence Conditions, Study 2. Error bars represent 1 SE. *** $p < .001$.

judgments about a group of agents who refused to engage in costly behavior. We used Cost (Major vs. Minor), Consequence (Major vs. Minor), Question Type (Lone, Single, Group; within-subjects) and Trial Order as predictors. We again included Age (in months) as a covariate and ID as a random effect in our model.

We began by comparing this model to a null model (with only Age, Trial Order, and Participant ID as predictors) Results from this comparison indicated that Cost, Consequence, and Question Type together explained a nonzero proportion of variation in children’s predictions, $\chi^2(4, N = 120) = 529.13, p < .001$. We found a main effect of Question type ($\chi^2(2, N = 120) = 317.63, p < .001$), a main effect of Consequence ($\chi^2(1, N = 120) = 323.85, p < .001$). There were no other significant predictors (all $ps > 0.484$). Children rated refusal to engage in costly behavior as less permissible when consequences were major ($M_{Major} = 0.89, SD = 1.07$) compared to when they were minor ($M_{Minor} = 1.98, SD = 0.84; t(579) = 20.67, p < .001$). They also rated refusal to engage in costly behavior as more permissible when a lone agent refused ($M_{LoneAgent} = 2.19, SD = 0.70$) compared to when a single actor in a group ($M_{SingleOfMany} = 1.16, SD = 1.08; t(597) = 15.85, p < .001$), and compared to the group as a whole ($M_{WholeGroup} = 0.95, SD = 0.107; t(597) = 19.07, p < .001$). Children’s responses did not differ when making judgments about a single member of the group versus the group as a whole ($t(597) = 3.22, p > .001$).

Finally, we compared their responses to the median possible score (1.5). For the Lone Agent question their responses were significantly greater than the median score in both cases ($M_{Minor} = 2.32, SD = 0.65; t(119) = 13.81, p < .001$), ($M_{Major} = 2.06, SD = 0.74; one-sample-t(119) = 8.30, p < .001$)). When enough agents refused to engage in costly behavior and they exceeded the threshold for collective consequence (Single of Group, Whole Group questions), children’s responses varied as a function of the consequences that were incurred. They rated refusal to engage in costly behavior as permissible when the collective consequence was minor [($M_{Minor, SingleofGroup} = 1.91, SD = 0.82; t(119) = 5.46, p < .001$), ($M_{Minor, WholeGroup} = 1.72, SD = 0.92; one-sample-t(89) = 2.58, p = .011$)]. On the other hand, they indicated that it was impermissible when there was a major collective consequence [($M_{Major, SingleofGroup} = 0.42, SD = 0.73; t(119) = -16.29, p < .001$), ($M_{Major, WholeGroup} = 0.19,$

$SD = 0.54; one-sample-t(119) = -26.58, p < .001$)). Together, these findings show that collective consequences, but not personal costs shape how children judge agents who refuse to engage in costly behavior for the greater good (See Fig. 5). Further, these results replicate the findings of Study 1 showing that hypothetical consequences are more permissible than realized ones, and that children hold individual group members collectively responsible for the consequences that occurred due to their refusal to engage in costly behavior.

3.4. Discussion

The findings of Study 2 provide further evidence that children are engaging in cost-benefit analysis when reasoning about agents who act for the greater good. We replicated the results of Study 1 in the extreme cases – children expect more agents to pay a minor personal cost to prevent a major collective consequence, and judge refusal to engage in costly behavior as less permissible. However, in Study 2 we demonstrated separate effects of costs and consequences on children’s reasoning about the greater good. In their predictions about how many people would act, children considered both factors. When costs were low, children expected a majority of individuals to engage in costly behavior, even to prevent minor consequences. When costs were high, however, children expected agents to engage in costly behavior more when consequences were major.

Children’s tendency to universalize, in contrast, was influenced by the severity of collective consequences only. In the case where one agent refused to engage in costly behavior and no consequences were realized, children universalized more (rated the action as less permissible) upon learning about the potential of major harms to the collective. Interestingly, they did so even when the personal cost of acting for the greater good was high, and despite the fact that they assumed most individuals would act selfishly. Of course, once harmful consequences were realized, children’s evaluations became more negative, but their evaluations (both of individuals and of groups) were still driven by consequences. The implications of these findings are discussed below.

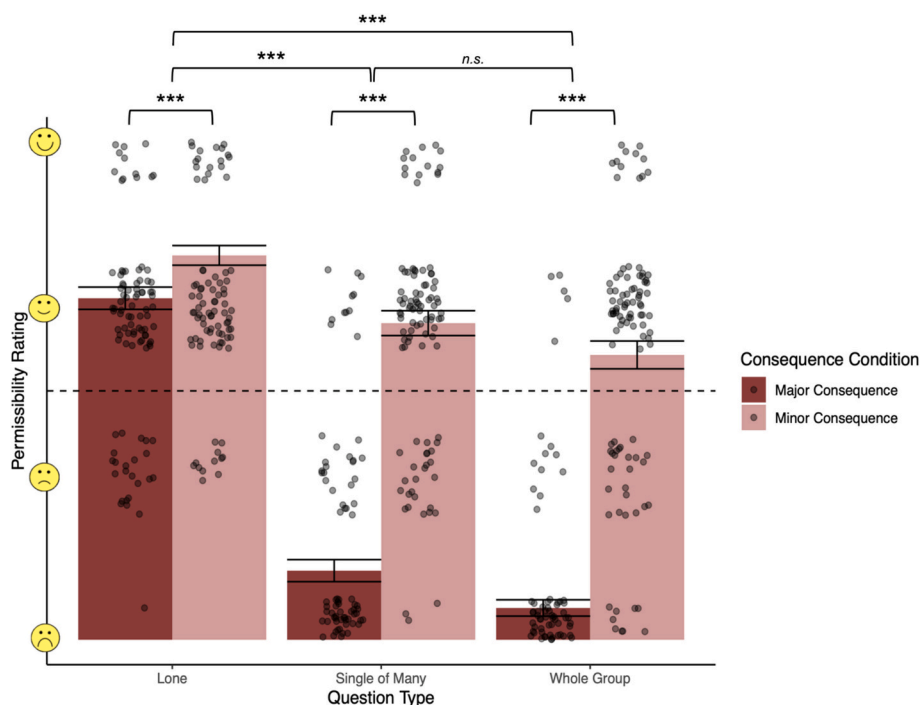


Fig. 5. Children’s responses to permissibility judgments by consequence condition, Study 2. Dashed line represents the mean possible permissibility rating, error bars represent 1 SE. *** $p < .001$. See supplement for responses by Cost and Consequence conditions.

4. General discussion

Acting for the greater good can be costly, but not all costs are worth paying. Instead, our expectations about when others should act for the greater good are informed by the tradeoff between the costs paid and the potential consequences incurred. The current study addressed how 6- to 9-year-old children use this information about tradeoffs when reasoning about the greater good. We explored this across three questions. First, we asked whether children predict interest (how many people) in acting for the greater good as a function of personal cost and collective consequence. Second, we asked how children's tendency to universalize is shaped by personal cost and collective consequence when individual actions are harmless and collective consequences are hypothetical. Finally, we asked how children make judgments about individuals and groups of self-interested actors after consequences have occurred.

Our findings showed that 6- to 9-year-olds' predictions about actions with collective benefits were shaped by a trade-off between personal cost and collective consequences – both in Study 1 (at the extremes) and in Study 2 (when personal cost and collective consequence were separated). These findings add to a growing literature showing how, across development, children use information about costs to guide their reasoning about others' actions (Aboody, Zhou, & Jara-Ettinger, 2021; Bridgers, Buchsbaum, Seiver, Griffiths, & Gopnik, 2016; Gergely & Csibra, 2003; Jara-Ettinger, Gweon, et al., 2015; Jara-Ettinger, Schulz, & Tenenbaum, 2020; Liu et al., 2017). Importantly these results confirm that, just like in cases of interpersonal needs (e.g., Radovanovic et al., 2023), children consider trade-offs when reasoning about collective needs, and further confirm anecdotal and empirical evidence that, at least by age 6, children readily link individual actions to societal welfare (e.g., Amemiya, Mortenson, Heyman, & Walker, 2023).

Perhaps most importantly, our findings shed light on the role of personal cost and collective consequence in children's tendency to universalize. At first, the results of Study 1 suggested a straightforward set of inferences whereby cost and consequences together influence the tendency to universalize, as well as influencing children's judgments of individuals and groups following the realization of collective consequences. However, in Study 2 we found that though children still weighed personal cost in their predictions, children's tendency to universalize – to base their judgments of a single harmless act on the hypothetical consequences that would occur if others acted similarly – was driven by the severity of the consequences, and not by the personal cost to carry them out.

Children's responses showed that they are universalizing when making judgments about a lone agent who refuses to engage in costly behavior for the greater good by considering the hypothetical consequences that could occur if others similarly refused – it was *less* okay to refuse when the hypothetical consequences were severe. However, children rated an individual refusal to engage in costly behavior as overall permissible regardless of the severity of the potential consequences. As a result, children rated refusal to engage in costly behavior as overall less permissible after consequences had occurred compared to when the consequences were hypothetical. But here again, children's judgments were based on the severity of the consequences to the group, regardless of how personally costly those actions were to carry out. This is further supported in children's justifications for their permissibility judgments – children overwhelmingly referenced consequences when asked why they made a particular judgment, and relatively few mentioned costs at all (See Supplement). Together, this suggests that collective consequence may be particularly influential when children are making judgments about agents who act for the greater good.

These results add nuance to the findings of Levine et al. (2020) as well as the findings of our Study 1: while children may consider self-interest (e.g., personal cost) to be a driver of behavior (as evidenced by their predictions), they privilege severity of consequences in their evaluations of such behavior (as evidenced by their permissibility judgments). This interesting tension suggests that at least children hold

individuals accountable for doing the right thing for the collective, even if it's difficult to do. On the face of it, these results also suggest important differences between individual moral actions and collective ones – prior work shows that when costly actions have consequences in and of themselves, adults and children consider the cost to the individual in their evaluations (e.g., Radovanovic et al., 2023; Starmans & Bloom, 2016; Zhao & Kushnir, 2022). But it may be the case that children don't consider individual costs relevant to evaluating actions as only consequential if many people engage in them (actions such as mask wearing, vaccination, conservation/recycling, etc.). Perhaps children are unique in this way, since we know from some prior work that adults are more likely to cite personal cost as a reason not to act for the greater good (e.g., Gidengil et al., 2019; Roy, Biswas, Islam, & Azam, 2022). Further research is needed to test out if differences between individual actions and collective actions bear out empirically in children, and also how this reasoning compares may change across the lifespan.

Another contrast between individual actions which have direct consequences and collective actions is that it involves thinking about collective responsibility. When a group of people refuse to act for the greater good, do children make different judgments about an individual versus the group as a whole? Our findings suggest that children do not make a distinction between individuals and groups in this case, rating both the individual and the group's actions as equally impermissible. One reason may be incidental; as we did not specify the order in which each agent dropped their umbrellas, it was not clear if children were thinking of the individual we pointed to as the last one, and thus the one responsible for exceeding the threshold. But another reason might be more substantive, that regardless of who did what when, children in our study were holding each individual collectively responsible for the consequences that occurred. This has support from the existing literature on how children believe blame and responsibility should be shared amongst group members. For instance, work showing that 5- to 7-year-old children believe responsibility for one's actions can be shared amongst group members (Over et al., 2016). This suggests that, at least when the individual cause for consequences is left ambiguous, children hold each individual equally responsible. Future work can explore whether this pattern remains when children are provided with information about whose individual actions actually led to the collective consequence occurring.

Several other open questions arise from our findings. First is the question of kinds of costs. Study 1 looks at how children think about two different kinds of costs – physical and psychological. While the overall pattern of results was consistent across types of cost, children overall expected more agents to pay a physical cost over a psychological one. Open questions remain about why this finding emerged. One possibility is that children's understanding of physical versus psychological costs changes throughout development. Previous work has shown that children make different inferences about physical and psychological costs, and that their understanding of different kinds of costs follows different developmental timelines (e.g., Radovanovic et al., 2023; Zhao & Kushnir, 2022).

Another possibility is that children's inferences were guided by different principles regarding action versus inaction. This has support in prior work where children rated actors who caused harm in moral dilemmas more negatively than an actor who caused harm through inaction (Hayashi, 2015; Powell et al., 2012). Further, this work demonstrated a developmental trend where older children (7- to 8-year-olds and 11- to 12-year-olds) and adults showed an omission bias (preference for harm caused by inaction) and younger children (5- to 6-year-olds; Powell et al., 2012) did not. Given that children do distinguish between action and inaction when reasoning about consequences, it is possible that similar reasoning guided their inferences about action versus inaction when reasoning about costs in Study 1. The age range in the current study is consistent with this developmental time frame such that we might expect to see a developmental trend emerge in our data; however, it is possible that, combined with the task demands, even older

children may still have struggled to understand psychological/inaction costs. Indeed, age alone was not a significant predictor of children's responses, and exploratory analyses revealed that it also did not interact with vignette condition (See Supplement) suggesting that children's understanding of different kinds of costs in this study did not increase with age. As such, open questions remain about how children's understanding of both psychological costs and consequences develop when reasoning about agents who act for the greater good.

Relatedly, the current work leaves an open question regarding developmental timelines more generally. The current study investigated this question with 6- to 9-year-old children. We did not anticipate any age effects to emerge in this range and indeed, across both studies and all test questions, age was not a significant factor. Given the evidence that universalization reasoning is present in younger children (Levine et al., 2020) and that younger children also engage in cost-benefit analysis about other kinds of moral problems (e.g., Jara-Ettinger, Gweon, et al., 2015; Jara-Ettinger et al., 2016), we hope to explore the developmental timeline of this reasoning by adapting our existing methodology to be suitable for younger age groups. On the other end, the current work leaves open questions about how adults and older children consider these tradeoffs in their predictions and third-party judgments. For example, previous work has shown that adults' political attitudes and social affiliations shape their own beliefs about which costs are worth paying (e.g., Coffey & Joseph, 2013; Doell, Pärnamets, Harris, Hackel, & Van Bavel, 2021; Hatemi, Crabtree, & Smith, 2019; Joslyn & Sylvester, 2019). As such, future work can investigate the considerations of the tradeoff between personal cost and collective benefit across the lifespan.

Culture, too, provides an avenue for future work. Children in our sample come from a fairly homogeneous background – the majority of our participants are white and come from liberal, wealthy, and educated families. However, cultural and socioeconomic background may shape the way that children consider tradeoffs between the individual and their community. For instance, current research has shown that children's contributions to household responsibilities vary as a function of their family's socioeconomic status (e.g., Alcalá, Rogoff, Mejía-Arauz, Coppens, & Dexter, 2014; Goodnow & Delaney, 1989). Given that socioeconomic status plays a role in how much children are paying costs (chores, for example) for the greater good (their household), we can expect that their third-party judgments may also be influenced by this factor. Relatedly, our sample was collected in the United States where, like many Western cultures, the individual is often prioritized over the collective (e.g., Killen & Wainryb, 2000; Markus & Kitayama, 1991; Triandis, Bontempo, Villareal, Asai, & Lucca, 1988). Previous work has shown interesting differences in children's social cognition across different cultures (e.g., Legare & Harris, 2016; Mert, Hou, & Wang, 2023; Wente et al., 2016; Zhao et al., 2021), including work specifically showing that children's use of cost information varies across cultures (Schäfer, Haun, & Tomasello, 2015). The current study asks children to consider what personal costs an individual may owe to their community. Given that this question pits the cost to the individual against the benefits to the collective, we might expect children from cultures that emphasize collective values to prioritize community at rates that are higher than we have demonstrated in our current sample.

Additionally, disagreement can arise across our individual perceptions of the magnitude of the personal costs and collective consequences. For instance, one may refrain from getting vaccinated because they believe the vaccine itself is harmful while others believe that same vaccine to be safe and effective. Likewise, one person may believe that a contagious disease is worth preventing, while others may not see this as a severe consequence at all. This difference in perceived costs is often what leads to different decisions about when one should act for the greater good. Levine et al. (2020) address this issue in terms of interested parties – the number of people who are interested in performing an act determines how widely one person's behavior is generalized. In the current study, children learned about a group of agents who were all seemingly in agreement about the magnitude of the costs and

consequences. Of interest is how children's responses may change if this scenario were to reflect individual differences in how we perceive costly actions for the greater good. If a group of individuals believe that a cost is too severe, this reduces the number of interested parties, thus limiting how broadly a certain act can be universalized. Future work can explore this further by asking how perceived costs and consequences may change how widely an act is universalized and consequently, how children judge those who refuse to engage in costly behavior for the greater good.

Finally, future research should consider how children reason about real-world instances where they themselves must act for the greater good. The current study presented children with a novel scenario, but children encounter this kind of question regularly in their daily lives (i. e., deciding whether to wait in line, raise their hand, or take only one piece from an unsupervised bowl of Halloween candy). Further, the COVID-19 pandemic presented children with a highly novel, salient, and high-stakes situation in which they need to engage in costly behavior (wearing an uncomfortable mask, getting a vaccine shot) to keep their community safe. Given its prevalence in young-children's lives, future work may investigate both how children think about these real-world cases and how children actually behave in these scenarios. Our study suggests that at least by age 6, appeals to the greater good (especially to severe collective consequences) might be a useful method of teaching children to accept and follow new norms.

As a final related point, we return to the original claims of Levine et al. (2020) that universalization explains how new societal norms arise out of our consideration of the greater good. We add our findings to this interesting theoretical claim and suggest that the severity of consequences for *not* acting for the greater good might play a central role in norm creation. Of course, societal norms are not created by 6-year-old children, but it is around this age that children do begin to create spontaneous "rules" games and other activities involving their peers (e. g., Göckeritz, Schmidt, & Tomasello, 2014; Köymen, Schmidt, Rost, Lieven, & Tomasello, 2015; Schmidt & Tomasello, 2012). Thus, it another empirical question is whether considerations of the greater good arise spontaneously in children's creation of peer norms, whether disagreements (e.g., Langenhoff, Srinivasan, & Engelmann, 2024) arise due to differences in children's beliefs about collective consequences and whether children ignore or consider individual difficulty when enforcing rules of games.

In conclusion, we provide evidence that by at least age 6, children consider both costs and consequences when thinking about the greater good. Children use information about the tradeoff between personal cost and collective consequence to determine when we are expected to act for the greater good. However, their judgments of those who do not pay these costs are primarily shaped by the consequences. Inaction that leads to a severe collective consequence is less permissible than the same inaction when stakes are lower. Their judgments are also guided by the logic of universalization – they consider the degree of collective consequence, even when that collective consequence has yet to occur. As such, these findings demonstrate that children can engage in complex reasoning about when we are expected to act for the greater good – they make principled judgments about which costs are worth paying.

CRediT authorship contribution statement

Zoe Finiasz: Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Montana Shore:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Formal analysis, Data curation. **Fei Xu:** Writing – review & editing, Methodology, Conceptualization. **Tamar Kushnir:** Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Data availability

I have included a link with the OSF page at the Attach File step

Acknowledgments

This research was supported by the National Science Foundation (SL-1955280) to T.K., the John Templeton Foundation Developing Belief Network sub-award to T.K and The Charles Lafitte Foundation. We thank Joanna Brooke, Umang Dhingra, Alyssa Gabbidon, Kyra Hoskin, and Justine Medveckus for help with data collection and coding. We also thank the staff of the Durham Museum of Life and Science for their partnership and ongoing support of our research. Finally, we thank the children and families in the Durham community who participated in this project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cognition.2024.106051>.

References

- Aboodo, R., Zhou, C., & Jara-Ettinger, J. (2021). In pursuit of knowledge: Preschoolers expect agents to weigh information gain and information cost when deciding whether to explore. *Child Development, 92*(5), 1919–1931.
- Alcalá, L., Rogoff, B., Mejía-Arauz, R., Coppens, A. D., & Dexter, A. L. (2014). Children's initiative in contributions to family work in indigenous-heritage and cosmopolitan communities in Mexico. *Human Development, 57*(2–3), 96–115.
- Amemiya, J., Mortenson, E., Heyman, G. D., & Walker, C. M. (2023). Thinking structurally: A cognitive framework for understanding how people attribute inequality to structural causes. *Perspectives on Psychological Science, 18*(2), 259–274.
- Bennett, M., & Sani, F. (2008). Children's subjective identification with social groups: A group-reference effect approach. *British Journal of Developmental Psychology, 26*(3), 381–387.
- Brewer, N. T., Chapman, G. B., Gibbons, F. X., Gerrard, M., McCaul, K. D., & Weinstein, N. D. (2007). Meta-analysis of the relationship between risk perception and health behavior: The example of vaccination. *Health Psychology, 26*(2), 136.
- Bridgers, S., Buchsbaum, D., Seiver, E., Griffiths, T. L., & Gopnik, A. (2016). Children's causal inferences from conflicting testimony and observations. *Developmental Psychology, 52*(1), 9–18.
- Bridgers, S., Jara-Ettinger, J., & Gweon, H. (2020). Young children consider the expected utility of others' learning to decide what to teach. *Nature Human Behaviour, 4*(2), 144–152.
- Coffey, D. J., & Joseph, P. H. (2013). A polarized environment: The effect of partisanship and ideological values on individual recycling and conservation behavior. *American Behavioral Scientist, 57*(1), 116–139.
- Doell, K. C., Pärnamets, P., Harris, E. A., Hackel, L. M., & Van Bavel, J. J. (2021). Understanding the effects of partisan identity on climate change. *Current Opinion in Behavioral Sciences, 42*, 54–59.
- Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences, 7*(7), 287–292.
- Gidengil, C., Chen, C., Parker, A. M., Nowak, S., & Matthews, L. (2019). Beliefs around childhood vaccines in the United States: A systematic review. *Vaccine, 37*(45), 6793–6802.
- Göckeritz, S., Schmidt, M. F., & Tomasello, M. (2014). Young children's creation and transmission of social norms. *Cognitive Development, 30*, 81–95.
- Goodnow, J. J., & Delaney, S. (1989). Children's household work: Task differences, styles of assignment, and links to family relationships. *Journal of Applied Developmental Psychology, 10*(2), 209–226.
- Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature, 450*(7169), 557–559.
- Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences, 108*(50), 19931–19936.
- Hatemi, P. K., Crabtree, C., & Smith, K. B. (2019). Ideology justifies morality: Political beliefs predict moral foundations. *American Journal of Political Science, 63*(4), 788–806. Hofstede, 2001.
- Hayashi, H. (2015). Omission bias and perceived intention in children and adults. *British Journal of Developmental Psychology, 33*(2), 237–251.
- Jara-Ettinger, J., Floyd, S., Tenenbaum, J. B., & Schulz, L. E. (2017). Children understand that agents maximize expected utilities. *Journal of Experimental Psychology: General, 146*(11), 1574–1585.
- Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naive utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences, 20*(8), 589–604.
- Jara-Ettinger, J., Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2015). Children's understanding of the costs and rewards underlying rational action. *Cognition, 140*, 14–23.
- Jara-Ettinger, J., Schulz, L. E., & Tenenbaum, J. B. (2020). The naive utility calculus as a unified, quantitative framework for action understanding. *Cognitive Psychology, 123*, Article 101334.
- Jara-Ettinger, J., Tenenbaum, J. B., & Schulz, L. E. (2015). Not so innocent: Toddlers' inferences about costs and culpability. *Psychological Science, 26*(5), 633–640.
- Josephs, M., Kushnir, T., Gräfenhain, M., & Rakoczy, H. (2016). Children protest moral and conventional violations more when they believe actions are freely chosen. *Journal of Experimental Child Psychology, 141*, 247–255.
- Joslyn, M. R., & Sylvester, S. M. (2019). The determinants and consequences of accurate beliefs about childhood vaccinations. *American Politics Research, 47*(3), 628–649.
- Karlsson, L. C., Soveri, A., Lewandowsky, S., Karlsson, L., Karlsson, H., Nolvli, S., & Antfolk, J. (2021). Fearing the disease or the vaccine: The case of COVID-19. *Personality and Individual Differences, 172*, Article 110590.
- Killen, M., Mulvey, K. L., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Morally-relevant theory of mind. *Cognition, 119*(2), 197–215.
- Killen, M., & Wainryb, C. (2000). Independence and interdependence in diverse cultural contexts. *New Directions for Child and Adolescent Development, 2000*(87), 5–21.
- Köymen, B., Schmidt, M. F., Rost, L., Lieven, E., & Tomasello, M. (2015). Teaching versus enforcing game rules in preschoolers' peer interactions. *Journal of Experimental Child Psychology, 135*, 93–101.
- Langenhoff, A. F., Srinivasan, M., & Engelmann, J. M. (2024). Disagreement reduces overconfidence and prompts exploration in young children. *Child Development, 95*(5), 1616–1627.
- Legare, C. H., & Harris, P. L. (2016). The ontogeny of cultural learning. *Child Development, 87*(3), 633–642.
- Levine, S., Kleiman-Weiner, M., Schulz, L., Tenenbaum, J., & Cushman, F. (2020). The logic of universalization guides moral judgment. *Proceedings of the National Academy of Sciences, 117*(42), 26158–26169.
- Liu, S., Ullman, T. D., Tenenbaum, J. B., & Spelke, E. S. (2017). Ten-month-old infants infer the value of goals from the costs of actions. *Science, 358*(6366), 1038–1041.
- Markus, H. R., & Kitayama, S. (1991). Cultural variation in the self-concept. In *The self: Interdisciplinary approaches* (pp. 18–48). New York, NY: Springer New York.
- Mert, N., Hou, Y., & Wang, Q. (2023). What lies ahead of us? Collective future thinking in Turkish, Chinese, and American adults. *Memory & Cognition, 51*(3), 773–790.
- Misch, A., Over, H., & Carpenter, M. (2016). I won't tell: Young children show loyalty to their group by keeping group secrets. *Journal of Experimental Child Psychology, 142*, 96–106.
- Over, H., Vaish, A., & Tomasello, M. (2016). Do young children accept responsibility for the negative actions of ingroup members? *Cognitive Development, 40*, 24–32.
- Powell, N. L., Derbyshire, S. W., & Guttentag, R. E. (2012). Biases in children's and adults' moral judgments. *Journal of Experimental Child Psychology, 113*(1), 186–193.
- Radovanovic, M., Chao, T. W. T., Onyshko, E., Huynh, Q. D. T., Liu, Y. L., & Sommerville, J. A. (2023). Not just if, but how much: Children and adults use cost and need to make evaluations about generosity across contexts. *Cognition, 238*, Article 105533.
- Roy, D. N., Biswas, M., Islam, E., & Azam, M. S. (2022). Potential factors influencing COVID-19 vaccine acceptance and hesitancy: A systematic review. *PLoS One, 17*(3), Article e0265496.
- Schäfer, M., Haun, D. B., & Tomasello, M. (2015). Fair is not fair everywhere. *Psychological Science, 26*(8), 1252–1260.
- Schmidt, M. F., & Tomasello, M. (2012). Young children enforce social norms. *Current Directions in Psychological Science, 21*(4), 232–236.
- Smith, C. E., & Warneken, F. (2016). Children's reasoning about distributive and retributive justice across development. *Developmental Psychology, 52*(4), 613.
- Starmans, C., & Bloom, P. (2016). When the spirit is willing, but the flesh is weak. *Psychological Science, 27*(11), 1498–1506.
- Thomas, S., Kelsey, C., & Vaish, A. (2024). No one is going to recess: How children evaluate collective and targeted punishment. *Social Development, e12730*.
- Triandis, H. C., Bontempo, R., Villareal, M. J., Asai, M., & Lucca, N. (1988). Individualism and collectivism: Cross-cultural perspectives on self-ingroup relationships. *Journal of Personality and Social Psychology, 54*(2), 323.
- Wente, A. O., Bridgers, S., Zhao, X., Seiver, E., Zhu, L., & Gopnik, A. (2016). How universal are free will beliefs? Cultural differences in Chinese and US 4- and 6-year-olds. *Child Development, 87*(3), 666–676.
- Wise, T., Zbozinek, T. D., Michelini, G., Hagan, C. C., & Mobbs, D. (2020). Changes in risk perception and self-reported protective behaviour during the first week of the COVID-19 pandemic in the United States. *Royal Society Open Science, 7*(9), Article 200742.
- Woo, B. M., Liu, S., Gweon, H., & Spelke, E. S. (2024). Toddlers prefer agents who help those facing harder tasks. *Open Mind, 8*, 483–499.
- Zhao, X., & Kushnir, T. (2022). When it's not easy to do the right thing: Developmental changes in understanding cost drive evaluations of moral praiseworthiness. *Developmental Science, 26*.
- Zhao, X., Wente, A., Flecha, M. F., Galvan, D. S., Gopnik, A., & Kushnir, T. (2021). Culture moderates the relationship between self-control ability and free will beliefs in childhood. *Cognition, 210*, Article 104609.