Original Articles

What's fair? How children assign reward to members of teams with differing causal structures

Karla Koskuba[a],[*], Tobias Gerstenberg[b], Hannah Gordon[a], David Lagnado[a], Anne Schlottmann[a]

[a] *University College London, United Kingdom*
[b] *Massachusetts Institute of Technology, United States*

ABSTRACT

How do children reward individual members of a team that has just won or lost a game? We know that from pre-school age, children consider agents' performance when allocating reward. Here we assess whether children can go further and appreciate performance in context: The same pattern of performance can contribute to a team outcome in different ways, depending on the underlying rule framework. Two experiments, with three age groups (4/5-year-olds, 6/7-year-olds, and adults), varied performance of team members, with the same performance patterns considered under three different game rules for winning or losing. These three rules created distinct underlying causal structures (additive, conjunctive, disjunctive), for how individual performance affected the overall team outcome. Even the youngest children differentiated between different game rules in their reward allocations. Rather than only rewarding individual performance, or whether the team won/lost, children were sensitive to the team structure and how players' performance contributed to the win/loss under each of the three game rules. Not only do young children consider it fair to allocate resources based on merit, but they are also sensitive to the causal structure of the situation which dictates how individual contributions combine to determine the team outcome.

1. Introduction

Deciding how to distribute resources fairly is of central importance in society. The question arises in all walks of life, from a teacher deciding how to reward a group of children after a successful project, to a manager distributing a bonus pot amongst her team. This study considers how young children distribute reward to members of a team who have just won or lost a game. We know already that young children can reward based on performance (Anderson & Butzin, 1978; Baumard, Mascaro & Chevalier, 2011; Melis, Altrichter, & Tomasello, 2013), contrary to classic views that they are egalitarian (Damon, 1977; Piaget, 1932). Here we study in two experiments whether children go beyond considering individual performance, and also take into account that what is a fair reward may depend on the causal role that someone's performance played in winning or losing.

1.1. Distributive justice in children

Traditional views on moral reasoning (Damon, 1977; Piaget, 1932) held that children into the early school years simply distribute resources equally, with 5-year-olds showing no sensitivity to contextual

information such as performance or need, preferring to share equally among three characters (Sigelman & Waitzman, 1991). A bias toward self-interest can further supersede principles of fairness when the children themselves stand to gain, with recent studies showing mostly egalitarian allocations, including allocations to themselves, only becoming predominant in children at around 6 to 7 years of age (Fehr, Bernhard, & Rockenbach, 2008; Rochat et al., 2009; Smith, Blake, & Harris, 2013). This tendency to distribute without regard to other potentially relevant factors such as individual merit, or need, was often interpreted to be the result of children's limited cognitive processing capacity and social perspective-taking (Damon, 1977; McGillicuddy-de Lisi, Watkins, & Vinchur, 1994; Sigelman & Waitzmann, 1991).

However, a growing body of literature has established that even young children have a sophisticated and differentiated sense for how resources are to be fairly distributed (Shaw, Choshen-Hillel, & Caruso, 2016). Melis et al. (2013) demonstrated that children's self-serving bias can be mitigated by equity considerations from early on: When sharing sweets with a puppet, children at 3 years were more likely to share equally if the puppet had helped them retrieve the sweets, but gave themselves more if the puppet had been unhelpful. Indeed, 21-months-olds are surprised if two characters are rewarded equally, when only

one has completed a chore (Sloane, Baillargeon, & Premack, 2012), demonstrating that toddlers already expect individuals to be rewarded according to their efforts. When asked to distribute cookies between two characters who put in different degrees of effort while baking them, 3-year-olds first gave one cookie each, but tend to give a third cookie to the character who put in more effort (Baumard et al., 2011). From 4 years onwards, children spontaneously assign rewards based on merit rather than in egalitarian fashion (Anderson & Butzin, 1978).

Children also consider factors other than merit: Children aged three to eight years, increasingly distribute resources according to characters' needs, when both are equally industrious (Rizzo & Killen, 2016). From age four, children can integrate need with merit when both vary (Anderson & Butzin, 1978), and from age 6 children balance multiple moral concerns more widely (Rizzo, Elenbaas, Cooley, & Killen, 2016).

In all these studies children distributed a fixed amount of resources between two characters. It is therefore not clear whether children considered both characters fully. Children could focus on rewarding one character's performance, and simply leave the remainder, without any consideration of the second character's performance. A less ambiguous approach is to let children allocate from two independent resource pots, one for each character. Preliminary work in our lab suggests that 4-year-olds can not only distribute fixed resources between two characters playing a collaborative game, but they also can reward these characters independently for their performance. This independent reward format is adopted for the research reported below.

The studies mentioned so far focus on rewards for positive actions, but one may also consider children's response to a negative act or outcome. This has not been studied as extensively, but in Hamlin, Wynn, Bloom, and Mahajan (2011) 19- to 23-month-old children were more likely to take a treat away from a badly behaved character rather than from the one who behaved well. Similarly, in Kenward and Dahl (2011) helpful puppets were allocated more biscuits than hinderers by 4 ½ year olds. It is unclear whether children actively punished the bad character, or focused on allocation to the meritorious character, but by 4 years children clearly consider intentionally harmful acts punishable, and more punishable than accidentally harmful acts (Cushman, Sheketoff, Wharton, & Carey, 2013; Leon, 1984). Five-year-olds also distribute unpleasant items as punishment to an anti-social adult (Kenward & Östh, 2015) and 6-year-olds punish unfair distributions even at a cost to themselves (McAuliffe, Jordan, & Warneken, 2015).

Taken together, these studies support a view that even pre-school children have a sense of equity and justice that includes merit-based allocation of resources; they can also take more than one factor into account in their distributive decisions. However, to our knowledge, previous research has not investigated whether young children penalize based on performance, similar to the merit-reward relationship. Our major focus, however, is whether children already see agents' performance in context, allocating reward/penalty to individuals not just based on their performance, but also based on how this performance contributes to group outcome.

1.2. Complex group structures and responsibility attribution in adults

Most studies discussed above considered the simplest situation: If effort is directly proportional to outcome, it makes intuitive sense to allocate reward proportionally as well. However, group dynamics are rarely so simple. Unlike effort, outcome may be categorical (winning or losing), not continuous. For instance, what if a given performance threshold needs to be reached for success, and going further beyond this threshold is irrelevant? A running team might qualify only if all members run a minimum time, or a quiz team might succeed as soon as one of its members answers a question correctly. In such cases is it still fair to reward proportionally to individual performance? Similar complications arise when performance is not satisfactory and the team loses. If all members of the running team are below threshold, are they all equally to blame, or does it matter how far an individual was from

the qualifying time? These examples highlight the different ways in which an individual may contribute to the group outcome, depending on the causal structure of the situation. Steiner's group functions (Steiner, 1972) formalize three common rules for three different types of causal structure. Under an additive rule, each member still contributes proportionally, up to the threshold for team success, for example, a group of children painting a classroom wall. Under a conjunctive rule, all members need to surpass a minimum threshold for success, for example, school playtime cannot start until each desk is clear. Under a disjunctive rule, only one member needs to surpass a threshold for success, for example, the whole team can win the math quiz if one member knows the correct answer.

It is not straightforward anymore what constitutes a fair allocation of reward when both individual effort, as well as the causal structure of the situation vary. For instance, one might think that one team member may be rewarded more than other team members, if this individual's performance is crucial for winning, as under a disjunctive rule. However, if all are above a given threshold as under a conjunctive rule, would an unequal reward based on differences in performance still be fair? We know from the studies discussed above that children reward based on performance. The present study looks at the extent to which children's reward allocations to individuals are influenced by the causal structures of the game, which determines how important the performance was for the team outcome.

To our knowledge, there has been no research to date exploring how differing causal structures influence children's reward attribution, but recent studies with adults (Gerstenberg & Lagnado, 2010; G&L hereafter; Lagnado, Gerstenberg, & Zultan, 2013; Zultan, Gerstenberg, & Lagnado, 2012) motivated the present work. G&L implemented the three causal structures discussed above as rules within a complex, speeded counting game common to all three structures. Participants counted triangles in a diagram, playing in a team with three other fictional players. Whether a particular round was won or lost depended on each player's accuracy and the team game rule (additive, conjunctive or disjunctive). After each round, participants were informed of the correct answer, with all players' counts shown. Participants then assigned responsibility to the players for the outcome of each round.

One finding was that the relationship between responsibility ratings and a player's performance differed between the three rules. In general, more inaccurate players were given less responsibility for winning and more for losing. This, however, was mediated by the causal structure: Under the disjunctive rule, in which the team won only if at least one player's count was 100% correct, players with small inaccuracies were given more responsibility for the loss and less for the win than under the other rules, where small inaccuracies could still contribute to a team win. Thus, adults considered players' individual performance in relation to the role it played for the team outcome. Here, we investigate how this sensitivity to the causal group structure develops, in two studies with a similar experimental paradigm to G&L, simplified to make it appropriate for children.

G&L's study suggested that adults use a type of counterfactual reasoning (Lewis, 1973) to determine how much responsibility each player deserved for the outcome. This involves imagining an alternative game scenario in which the player arrives at a different count; if this imagined count changes the outcome from actual win to imagined loss (or vice versa), then the player is deemed responsible for the actual outcome. G&L found that participants' attributions were well accounted for by Chockler and Halpern's (2004) modified counterfactual model (also called structural model) in which responsibility depends on how close a player's contribution was to having made a difference to the team outcome.

Children's counterfactual reasoning increases over the age range studied here. While children from age 3 or 4 can consider counterfactual states in very simple circumstances, more complex considerations of what might have been, involving detailed consideration of differences/commonalities between the imagined and real world, as

done by G&L's adults, develops only slowly (Beck & Riggs, 2014; Beck, Robinson, Carroll, & Apperly, 2006; Rafetseder, Cristi-Vargas, & Perner, 2010.)

In Experiment 1, we address whether children differentiate the three game rules. Then in Experiment 2, we use modeling to explore whether children begin to use counterfactual reasoning similar to adults.

1.3. Overview of experimental paradigm

This study explores the sensitivity of children's merit allocations to different causal functions. Because responsibility attribution per se may be too abstract for young children, we consider reward allocation, in keeping with the literature reviewed above. We study whether children from age 4 or 5, who are capable of merit-based, proportional reward distribution (Anderson & Butzin, 1978; Baumard et al., 2011), also consider more complex causal structures, where individual performance is not always proportional to group outcome. The group outcome can be positive or negative, winning or losing a game, so a second issue is whether children capable of merit-based, positive reward allocation are also capable of similar penalty allocation.

Rather than G&L's counting game, we employed a setup in which characters throw balls at a target board with three concentric rings (similar to darts, see Fig. 1). This setup is more appropriate and engaging for children. The three causal functions (additive, conjunctive, disjunctive) were implemented as rules within the context of a two-character team playing the game. Under the additive rule, the team needed to get at least four stars to win, with three stars awarded for hitting the center region, two for the middle ring, one for the outer ring. Under the conjunctive rule, both team characters needed to get their ball in either the center or middle ring to win, visually represented on screen by the 'losing' outer ring colored red. Under the disjunctive rule, the team won if at least one player hit the center region, visually represented by the 'winning' center circle colored red.

In the four phases of the experiment, children first threw balls at the board themselves (Fig. 1a), then they learned, with a visual aid, one of the causal rules for winning (b). Over the next rounds, to consolidate rule learning, children predicted win/loss outcome from seeing the results of players' throws (c). Finally, during the experimental trials (d), children awarded sweets to each player on each of multiple game rounds. These rounds included all possible combinations of the two players' performance, from both hitting the center (CC) to both hitting the outer ring (OO). For some combinations, all three rules imply the same outcome (e.g., CC always wins), but for other combinations, the rules lead to different outcomes (e.g., CO wins under the disjunctive and additive rules, but loses under the conjunctive rule.)

We studied 4 to 5-year-old and 6 to 7-year-old children, as well as adults, included to assess whether any disparities with G&L's study reflect developmental differences or task differences. For adults, we

expected a pattern similar to that in G&L. It is not clear, however, as discussed above, that young children can engage in the counterfactual reasoning necessary for this, so our study was exploratory with respect to specific age differences.

Three possible patterns of results were envisaged: One is that children do not consider the different game rules, but assign sweets in proportion to individual performance. The latter has already been demonstrated in the studies reviewed above, and we hope to extend the finding from rewards for winning to penalties for losing. If children consider only individual performance, then players of balls landing in the center circle would gain most sweets in winning games or lose least in losing games, and players of balls landing in the outer circle would receive least or lose most sweets, with no effects of the different game rule conditions.

An alternative pattern of results is that children do consider the different causal structures, rewarding wins and penalizing losses, but ignore individual performance. This would mean that players with different levels of performance receive the same level of reward if they are part of winning teams, and the same level of penalty if they are part of losing teams. For instance, under the additive rule, an individual may hit the center, middle or outer ring and still be part of a winning team if the other player's performance compensates. In this case, both players will get full reward, regardless of individual performance.

Finally, if participants evaluate players' performance relative to the game rules, then both individual performance and whether the team won or lost under the given rule will be considered, with different patterns of reward allocations expected for identical patterns in performance, under different rules. If children behave like adults in G&L, then we would expect that reward for those individuals who did not contribute to a win (e.g., hitting the middle ring, rather than the center circle, under the disjunctive rule) would be less than under rules, in which the same performance level does contribute to the win.

2. Experiment 1

2.1. Material and methods

2.1.1. Participants

A total of 119 participants took part. This included 41 reception class children, in their first year of UK compulsory schooling – average age 5 years 0 months (range 4 years 3 months to 5 years 8 months, 22 girls), and 44 children from the class two years above – average age 6 years 11 months (range 6 years 4 months to 7 years 8 months, 24 girls). None of the children had special needs. The children took part in the study at the permission of their parents and were all from state primary schools in Greater London and Berkshire. The 34 adult participants (average age 33 years, range 18 years to 67 years, 16 females) were an opportunity sample. All materials including data, experiments,

a)	b)	c)	d)

Physical game playing to become familiar with the task.

Illustration of the rule during explanation ('additive' example).

Children watch animated throw and predict whether team won or lost, by clicking on the corresponding trophy. Correct or incorrect clicks are followed by a voice saying "right" or "wrong".

Children see the outcome of each trial on the board. Ball colour matches T-shirt colour of the thrower. Children respond by moving the sliders up or down, increasing or decreasing the size of the corresponding sweet bag.
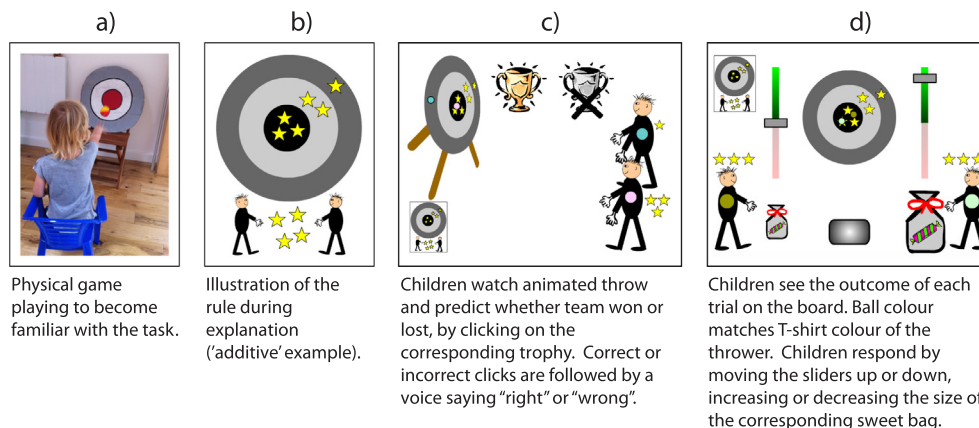
Fig. 1. The four experimental phases.

and analysis scripts are available at https://osf.io/4s6vb/ (Koskuba & Gerstenberg, 2018).

2.1.2. Materials and procedure

Children were tested individually in one session of about 20 min in a quiet room in their school. Initially, each child was familiarized with the physical game, involving a dart board with three concentric rings: a red center circle (C), white middle ring (M) and grey outer ring (O), each ring 10 cm thick, and two soft balls. Children tried 6–8 times, to hit the center circle of the board, placed at a distance of about 1.5 m, such that children (and adults) were not always successful, but managed to hit the center occasionally (see Fig. 1a).

The same set up was subsequently shown implemented on the computer, with two characters throwing one ball each at the dart board. Each character's t-shirt and ball had the same color so that children could easily identify who threw which ball. Children were told that the two characters were playing a game together, and could win under the allocated game rule, which was explained with a visual representation (see Fig. 1b). The rule explanation was illustrated with computer-animated examples of winning and losing game rounds. As a memory aid, the visual rule depiction was present in the left of the screen throughout the game (see Fig. 1c and d: Additive example).

To establish rule comprehension, children watched several animated throws and predicted the win or loss outcome by clicking on icons of a gold trophy or grey crossed out trophy (Fig. 1c). After each prediction, they received visual and auditory feedback (a green screen flash and a computer voice saying 'right' or a red screen flash and 'wrong'). When a child had given at least five correct responses in six consecutive trials, the child moved on to the experimental reward allocation trials.

On the first reward allocation trial, the display changed from an animation of the throw itself to a static display of the two players' performance pattern (ball positions on the board), with the two players, bags of sweets, and sliding scales for independent reward allocation, shown to each side (Fig. 1d). Stars above the players' heads provided an additional reminder of individual performance levels. On each trial, children also received an auditory cue indicating the team status, with a cheer for winning teams and a "duh" sound for losing teams. The slider was a bar positioned in the middle of the scale at the beginning of each trial. Linked to each slider was an endowment bag of sweets, initially of medium size, that increased in size as the slider was moved up to reward (coded 0 to 10), or decreased in size when it was moved down to punish (coded 0 to −10). The slider was constrained such that it could only slide up (increasing sweets) for winning games, and down for losing games. Each game round consisted of both characters throwing a single ball at the board at which point the game was either won or lost, depending on the game rule assigned. Participants then decided what size of sweet bag to assign to each team character; they had information on each character's performance as well as the team outcome (win or lose), both of which could be used to decide on the reward size. On the next trial, the slider was repositioned in the middle and the participant saw a new team, with different players indicated by novel t-shirt and ball colors, so that team history would not contaminate children's allocations.

Children and adults typically moved the slider themselves, but a few younger children had trouble controlling the mouse. These children pointed at the screen and the experimenter followed with the mouse. For adult participants, the initial throwing of physical balls at the circular board was omitted.

2.1.3. Design

For each child, the design was a 3 (left player performance: center region, middle ring, outer ring) × 3 (right player performance) factorial, yielding nine possible performance combinations (ball configurations on the board, OO, OM, MO, OC, MM, CO, MC, CM, CC). The additive, disjunctive and conjunctive rules result in 6, 5, and 4 winning trials respectively in these nine combinations, and three filler trials were added to each rule condition, so that all involved six wins and six losses. These filler trials were randomly drawn from winning or losing combinations, as needed, and were not included in the final analysis. The total of 12 trials in each condition was repeated twice, with the 24 trials randomized separately for each child. The overall design was a 3 (age) × 3 (game rule) × 3 (left player performance) × 3 (right player performance) design, with age and game rule varied between participants and left and right player performance varied within participants. Children were randomly assigned to one of the game rules, with 41, 39, and 39 participants in the additive, conjunctive and disjunctive rules, respectively.

3. Results

Participants assigned sweets to both characters within the team, therefore each trial produced two data points, one with the left player as the target, and one with the right player as the target. With nine team performance combinations, this created 18 separate data points. Preliminary analyses established that whether the target was on the left or right of the screen did not affect the results, e.g., reward to the left player for performance combination CM (left/right) corresponded to reward to the right player for combination MC (left/right). Allocations for equivalent performance combinations of left and right targets were therefore averaged, leaving nine data points. The labels CC, CM, MC, etc., now reference the target first, the other player's performance second, and not their position on screen. Further analyses established that the gender of the participant, and the experimenter running the sessions had only minor effects on reward allocation; these factors are not considered further.

Fig. 2 shows the main results (Koskuba & Gerstenberg, 2018), how three different ages (top to bottom panels) allocated sweets under the three different winning rules (separate curves) to a target character as a function of the nine different performance combinations (horizontal). The curves generally increase towards the right, at all ages. More sweets were awarded for better play or taken away for worse play. As in previous work (Anderson & Butzin, 1978; Baumard et al., 2011; Sloane
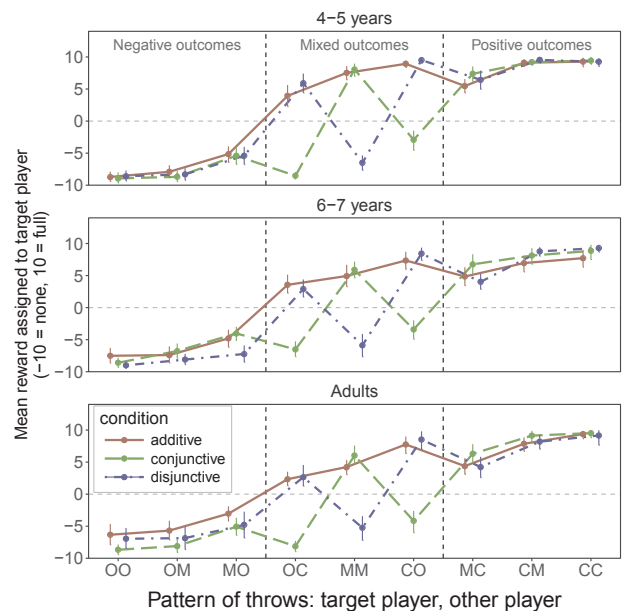
Fig. 2. Mean reward assigned to target player in Experiment 1 across the nine performance combinations for different age ranges (panels) and rules (lines). The first label on the x-axis refers to the performance of the target player, and the second label refers to the performance of the other player. O = outside ring, M = middle ring, C = center ring. For example, "CM" indicates that the target player hit the center, and the other player hit the middle ring. *Note*: Error bars indicate 95% bootstrapped confidence intervals.

et al., 2012), even the youngest children allocated reward in line with performance.

Of main interest is whether children consider the different game rules in their reward allocation, reflected in differences between the curves on each panel; to aid this analysis, the graphs are divided into three sections. The right section (MC, CM, CC) shows performance combinations that win under all three game rules. The left section (OO, OM, MO) shows combinations that lose under all rules. The middle section (OC, MM, CO) shows combinations that can win or lose, depending on the rule. For instance, CO wins under the disjunctive (at least one character got the bulls eye) and additive rule (together, players got four stars), but loses under the conjunctive rule (both players needed to hit at least the middle ring, but the non-target missed). Rule differentiation is clear in this midsection, but could simply reflect the game set-up, since Experiment 1 was constrained such that if the team won, children could only award sweets, not take them away, and vice versa if the team lost. The subtler curve separation and potential rule differentiation in the left and right sections of each panel are of more interest since they are not forced by the game set-up, reflecting children's unhindered decision making. We therefore focus on how participants attributed rewards as a function of the different rules for the 'all rules win' and 'all rules lose' trials.

3.1. Rule impact on Performance: 'all rules win' trials

The overall upward trend in the right section of each panel for winning games (MC, CM, CC), was reflected in a main effect of performance $F(1.55,170) = 175$, $p < 0.001$, partial $\eta^2 = 0.61$, in a 3 (*performance combination*) × 3 (*rule*) × 3 (*age group*) ANOVA, with repeated measures on performance combinations. Thus, children rewarded based on players' merit.

More importantly, the performance × rule interaction $F(3.10,170) = 5.06$, $p = 0.002$, partial $\eta^2 = 0.084$, suggests that children reward the same performance differently under different rules, but the visual effect (right of Fig. 2), is subtle when presented on the full allocation scale. The clearest pattern appears at 6–7 years, with the slope for the disjunctive rule from MC up to CM steeper than for additive or conjunctive rules: The reward given to the target player in the MC combination increases more under the disjunctive rule. This makes sense: Only one player needs to hit the center to win, so the target alone produces the win in the CM combination, but does not contribute to it with MC. Under the other rules, in contrast, the target contributes directly to the win for both MC and CM. The graphical pattern was similar for adults, but not clear for 4–5-year-olds. Nevertheless, the rule × performance interaction did not differ between the ages $F(6.20, 170) = 1.58$, $p = 0.15$, partial $\eta^2 = 0.054$.

There was also a main effect of rule, $F(2,110) = 4$, $p = 0.020$, partial $\eta^2 = 0.068$, with children awarding the least sweets under the additive, and the most sweets under the conjunctive rule, and of age $F(2,110) = 4.53$, $p = 0.013$, partial $\eta^2 = 0.076$, with 4–5-year-olds awarding slightly more sweets than older participants. No other effects were significant (performance × age: $F(3.10,170) = 2.32$, $p = 0.075$, partial $\eta^2 = 0.040$, age × rule: $F < 1$).

When the ages were analyzed separately, all ages showed a main effect of performance combination (4–5-year-olds: $F(1.31,49.9) = 47.7$, $p < 0.001$, partial $\eta^2 = 0.56$, 6–7-year-olds: $F(1.69,69.4) = 62.6$, $p < 0.001$, partial $\eta^2 = 0.60$, and adults: $F(1.50,46.6) = 63.8$, $p < 0.001$, partial $\eta^2 = 0.67$), confirming that children as young as 4 years reward based on merit. Only the older children showed a significant performance × rule interaction, $F(3.39,69.4) = 6.19$, $p = 0.001$, partial $\eta^2 = 0.23$. No other effects were significant (4–5-year-olds: rule, $F(2,38) = 1.54$, $p = 0.23$, partial $\eta^2 = 0.075$; performance × rule, $F(2.63,49.9) = 1.67$, $p = 0.19$, partial $\eta^2 = 0.081$; 6–7-year olds: rule, $F(2,41) = 1.62$, $p = 0.21$, partial $\eta^2 = 0.073$; Adults: rule $F(2,31) = 1.73$, $p = 0.19$, partial $\eta^2 = 0.10$, performance × rule $F(3.01,46.6) = 1.05$, $p = 0.38$, partial $\eta^2 = 0.063$).

3.2. Rule impact on performance: 'all rules lose' games

On the left of Fig. 2, for performance combinations OO, OM, MO, that lose under all rules, the older two ages appear to differentiate between rules, but this did not manifest in the analysis, with no effects for game rule. There were only main effects of performance, $F(1.64,181) = 117$, $p < 0.001$, partial $\eta^2 = 0.52$, seen in the increase in reward across the three combinations, and of age $F(2,110) = 4.26$, $p = 0.017$, partial $\eta^2 = 0.072$, with 4–5-year-olds penalizing slightly more harshly than other ages. No other effects were significant (performance × age $F < 1$, performance × rule $F(3.28,181) = 2.13$, $p = 0.092$, partial $\eta^2 = 0.037$, performance × rule × age $F < 1$, rule $F(2,110) = 3.02$, $p = 0.053$, partial $\eta^2 = 0.52$, rule × age $F(4,110) = 2.25$, $p = 0.068$, partial $\eta^2 = 0.076$).

3.3. 'Defiant' responses

An incidental, but potentially relevant occurrence was that children, particularly 4–5-year-olds, tried to reward against the constraints. For instance, this happened for performance combinations such as MC and OC under the disjunctive rule; here the team wins, but children often tried to reduce the target's number of sweets, presumably because the target did not contribute to the win. Similarly, children wanted to reward the target player when the team lost but the target did not contribute to the loss, in performance combinations such as CO and MO under the conjunctive rule. The program constraints did not allow this and when children realized this, some kept the target at the initial state (not rewarding or penalizing), and, perhaps in compensation, gave or took the full amount from the other player. These categorical response tendencies may well have contributed to the age main effects reported above. Other children, however, seemed confused when they could not allocate sweets against the constraint and ended up with seemingly nonsensical responses.

Children's 'defiant' responses in Experiment 1 were unexpected, so we cannot quantify them. They were obvious and frequent, however, especially in the youngest age group, and motivated our second experiment below.

4. Discussion

Overall, Experiment 1 showed that participants considered performance when allocating rewards, as well as penalties, mirroring and extending previous work (Anderson & Butzin, 1978; Baumard et al., 2011; Melis et al., 2013). With respect to sensitivity to causal group structures, results were ambiguous: for winning games, rewards differed by rule. However, rule differentiation did not appear for all individual age groups, or for losing games.

Unexpectedly, children tried to reward players in a losing team or penalize in a winning team, possibly focusing on individual performance rather than group outcome. Such attempts were most apparent for 4–5-year-olds, but the constraint on responding potentially also affected 6–7-year-old children; it may have scaffolded answers by forcing a consideration of group outcome rather than individual performance, supporting advanced rule differentiation.

The logic behind the procedure of Experiment 1 was that wins would naturally lead to increases in reward, losses to decreases, i.e., penalties, mirroring G&L's responsibility allocation study. Young children's 'defiant' behaviour suggests that they did not share this intuition and that reward and responsibility for an outcome can come apart (cf. Cushman, 2008). Attempts to reward a loss and penalize a win could reflect that children find individual effort more salient than the group outcome. This may be because they did not learn the game rules for winning and losing as well as older participants, or they forgot the rule during the test phase. These points were addressed in Experiment 2.

5. Experiment 2

Experiment 2 assessed children's natural reward allocation tendencies by removing the constraint on responding. Participants were able to make use of the full scale, which means that they were allowed to reward and penalize in both winning and losing games. Adults were expected to use the previous strategy of rewarding wins and penalizing losses, mediated by performance. Children were expected to revert more to rewarding individual performance regardless of rule. Increased age differences were expected, with potentially less rule differentiation at younger ages.

In Experiment 2, we also model children's reward allocations, to gain better insight into the underlying cognitive processes. We will fit the models previously considered in G&L's work on group functions in adults' responsibility attribution, as well as other models more specifically suited to reward allocation. We fit these models at the individual level. The behavioural data provide evidence on rule differentiation at the group level, but the modeling may allow us to see whether individuals use different strategies under different rules.

5.1. Method

5.1.1. Participants

There were 135 participants in this experiment. Of these, 45 were children from the first year of school, average age 4 years 10 months (range 4 years 1 month to 5 years 8 months), and 45 were children from the class two years above, average age 6 years 9 months (range 6 years to 7 years 5 months). There were 46 girls, 44 boys and none had any special needs. The children took part in the study at the permission of their parents and were all from state primary schools in Berkshire. The 45 adult participants, average age 27 years (range 19 to 74 years) were an opportunity sample.

5.1.2. Procedure

The apparatus, design and procedure were similar to Experiment 1. The only difference from the participant's perspective is that the sliders were unconstrained, so participants could remove sweets from characters in winning teams and increase sweets to characters in losing teams. As in Experiment 1, the slider's initial position was in the middle of the scale, giving characters a medium sized bag from which participants could add or remove sweets. The only other difference was that for each participant we tracked the number of correct and incorrect outcome predictions prior to reward allocation, in order to evaluate potential age differences in rule learning.

6. Results

6.1. Rule learning

The 4–5-year-olds, 6–7-year-olds, and adults needed 5.8, 5.4, and 5.1 trials, respectively, to reach our criterion for rule learning (5 correct outcome predictions in 6 consecutive trials), a significant decrease with age, $F(2,128) = 4.18$, $p = 0.017$, partial $\eta^2 = 0.061$. Only 5 children needed more than 7 trials. In absolute terms, the game rules were easy to learn, at all ages.

6.2. Rule differentiation

Children's mean reward allocations in Experiment 2 are in Fig. 3. Overall the data look similar to Experiment 1, with an overall increase in reward as performance improves. As previously, each panel has three sections, for games that lose (left) or win (right) under all rules, and for games with different outcomes under different rules (middle). The midsection shows clearest separation of curves for the three rules, but this time it is not forced by the procedure. It reflects children's unconstrained reward allocations and indicates a real understanding of
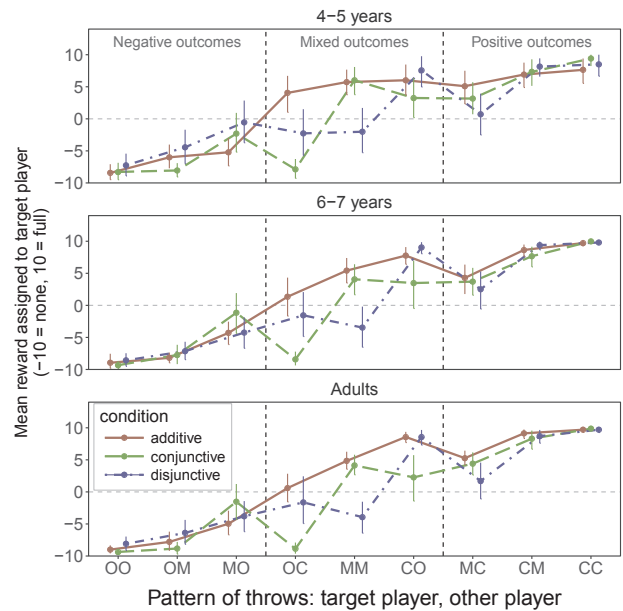
Fig. 3. Mean reward assigned to target player in Experiment 2 across the nine performance combinations for different age groups (panels) and game rules (lines) show that children of all ages reward differently under the three rules, for both winning and losing combinations. The first label on the x-axis refers to the performance of the target player, and the second label refers to the performance of the other player. O = outside ring, M = middle ring, C = center ring. For example, "CM" indicates that the target player hit the center, and the other player hit the middle. *Note*: Error bars indicate 95% bootstrapped confidence intervals.

the rule differences. The left and right sections also show subtle rule differences, contrary to our expectation of less rule differentiation in Experiment 2.

6.2.1. Rule impact on performance: 'some rules win, some lose' games

The midsection of Fig. 3 shows clear visual differentiation for performance combinations (OC, MM, CO) that win under one rule, but lose under another. For instance, MM wins under conjunctive and additive rules, and sweets are awarded, but loses under the disjunctive rule and sweets are taken away. Similar rule differentiation appears for CO and OC, and in all three age groups: Even the youngest children reward not just individual performance, but consider this in the context of the causal game structure.

A 3 (*performance combination*: OC, MM, CO) × 3 (*rule*) × 3 (*age group*) ANOVA with repeated measures on performance combinations reveals a main effect of rule, $F(2,126) = 32.7$, $p < 0.001$, partial $\eta^2 = 0.34$, of performance, $F(2,252) = 113$, $p < 0.001$, partial $\eta^2 = 0.47$, and an performance × rule interaction, $F(4,252) = 33.5$, $p < 0.001$, partial $\eta^2 = 0.35$, since each rule wins a different numbers of times (additive for all, OC, MM and CO, disjunctive for OC, CO and conjunctive only for MM). There were no age differences. (Age $F < 1$; rule × age $F < 1$; performance × age $F(4,252) = 1.04$, $p = 0.39$, partial $\eta^2 = 0.02$; performance × rule × age $F < 1$)

The average allocated to CO under the losing conjunctive rule, while less than for the winning disjunctive and additive rules, is above 0 for all age groups, indicating that participants increased sweets to a character in a losing team. Similarly, OC wins under the disjunctive rule, but the average allocated is below 0, indicating that participants sometimes took sweets away from the target in a winning game. This highlights a tendency to respond against the group outcome, as noted in Experiment 1, and we return to this issue below.

6.2.2. Rule impact on performance: 'all rules win' games

For the winning panel, the same pattern is seen as in Experiment 1,

with steeper slope from MC to CM for disjunctive (dashed line) than the other rules, for all three age groups. The 3 (*performance combination*) × 3 (*rule*) × 3 (*age group*) ANOVA for the winning combinations (MC, CM, CC) found a main effect of performance combination, $F(1.43,181) = 149$, $p < 0.001$, partial $\eta^2 = 0.54$, and more importantly, a significant interaction of performance combination with rule, $F(2.87,181) = 6.43$, $p < 0.001$, partial $\eta^2 = 0.093$. There were again no age differences. (Age $F(2,126) = 2.54$, $p = 0.08$, partial $\eta^2 = 0.04$; performance $F(2,126) = 1.16$, $p = 0.32$, partial $\eta^2 = 0.02$; performance × age and performance × rule × age, $F < 1$)

When the three ages were analyzed separately, all showed a performance effect (4–5-year-olds: $F(1.57, 66.1) = 33.8$, $p < 0.001$, partial $\eta^2 = 0.45$, 6–7-year-olds: $F(1.32, 55.6) = 55.3$, $p < 0.001$, partial $\eta^2 = 0.57$, adults: $F(1.34, 56.4) = 69.5$, $p < 0.001$, partial $\eta^2 = 0.62$).); the rule effect was not significant (4–5-year-olds and 6–7-year olds: $F < 1$; Adults: $F(2,42) = 1.92$, $p = 0.16$, partial $\eta^2 = 0.08$). The rule × performance interaction was significant for 4–5-year-olds, $F(3.15, 66.1) = 3.62$, $p = 0.016$, partial $\eta^2 = 0.15$), marginal for adults, $F(2.69, 56.4) = 2.61$, $p = 0.067$, partial $\eta^2 = 0.11$, and not significant for 6–7-year-olds, $F(2.65, 55.6) = 1.14$, $p = 0.34$, partial $\eta^2 = 0.051$ Nevertheless, the MC to CM slope is steepest under the disjunctive rule at all ages, in line with the absence of age differences in the more powerful overall analysis.

6.2.3. Rule impact on performance: 'all rules lose' games

A complementary pattern appeared for losing performance combinations OO, OM, MO. Rule differentiation was clearer than in Experiment 1, with more sweets given to the target character in the MO than OM and OO combinations under the conjunctive rule than the other two rules, i.e., the slope of the dotted line is steeper. This makes sense: Under the conjunctive rule (both must hit the two middle rings), the target character does not contribute to the loss, in contrast to the OM and OO combinations. For disjunctive and additive rules, the target contributes to the loss in all performance combinations. This rule differentiation is reflected in an interaction of performance combination with rule, $F(3.06,193) = 6.13$, $p < 0.001$, partial $\eta^2 = 0.089$ in the 3 (*performance combination*: OO, OM, MO) × 3 (*rule*) × 3 (*age group*) ANOVA. A significant main effect of performance combination, $F(1.53,193) = 111$, $p < 0.001$, partial $\eta^2 = 0.47$ again reflects the reduction in sweets as target performance worsens. There were no other significant effects, in particular, no age differences (performance × age $F < 1$; performance × age × rule $F(6.12,193) = 1.00$, $p = 0.43$, partial $\eta^2 = 0.03$; age $F(2,126) = 1.81$, $p = 0.17$, partial $\eta^2 = 0.03$; rule $F(2,126) = 2.50$, $p = 0.09$, partial $\eta^2 = 0.04$; age × rule $F < 1$).

When the age groups were analyzed separately, all showed main effects of performance (4–5-year-olds: $F(1.64, 69.0) = 26.8$, $p < 0.001$, partial $\eta^2 = 0.39$, 6–7-year-olds: $F(1.52, 63.8) = 37.5$, $p < 0.001$, partial $\eta^2 = 0.47$, adults: $F(1.30, 54.7) = 56.8$, $p < 0.001$, partial $\eta^2 = 0.58$). The rule × performance interaction only reached significance for adults ($F(3.29, 69.0) = 2.42$, $p = 0.069$, partial $\eta^2 = 0.10$, $F(3.04, 63.8) = 1.93$, $p = 0.13$, partial $\eta^2 = 0.084$, $F(2.60, 54.7) = 4.64$, $p = 0.008$, partial $\eta^2 = 0.18$ for 4–5-year-olds, 6–7-year-olds and adults, respectively). The rule main effect was not significant. (4–5-year olds: $F(2, 42) = 2.19$, $p = 0.12$, partial $\eta^2 = 0.10$; 6–7-year olds: $F < 1$; adults $F < 1$.)

Overall, and contrary to expectation, children in Experiment 2 showed clearer rule discrimination than in Experiment 1. Note in particular that the steeper slope under the conjunctive rule for the 'all rules lose' sections combined with steeper slope under the disjunctive rule for the 'all rules win' section rules out a simple form of rule differentiation, such that children generally put more weight on the performance dimension in one rule condition than another. Instead, an account is required which is sensitive to how performance and rule interact: Even the youngest children's reward allocation reflected that the same level of performance can contribute to the outcome differently, depending on the causal game rule.
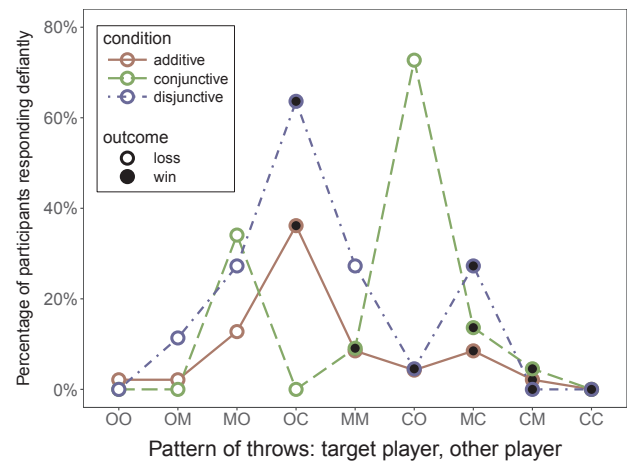

Fig. 4. Percentage of defiant participants (who decreased reward for a win, or increased reward for a loss) across nine performance combination, separated by rule condition. The circle fill indicates whether the team lost or won depending on the performance combination and rule. For example, when the target player hit the center, and the other player hit the outside ring (CO), 75% of participants increased the reward for the target player even though the team lost in the conjunctive rule condition.

6.3. 'Defiant' responses

In Experiment 2, participants could reward and penalize regardless of whether the team won or lost, and not only children, but adults did so. The percentage of participants for each performance combination who reduced reward for wins or increased for losses is in Fig. 4.

In Fig. 4, responses contrary to the group outcome appear more often when the target's performance did not fit with the team result under a given rule. For the CO performance combination, for example, although the target C performs well, under the conjunctive rule the team loses. Nevertheless, over 70% rewarded the target C. Similarly, for OC, the team wins under the disjunctive rule, but the target does not contribute, and over 60% penalized the target. With smaller discrepancies, for instance, for MC under the disjunctive rule, defiant responses are less frequent, but they appear for all performance combinations except CC.

We had expected such responses for younger children, but they appeared at all ages. This was confirmed in three-way loglinear analyses (between age group, rule and whether a response was 'defiant') for each combination, which found only a single significant age effect across the nine analyses: The youngest age was more likely to produce defiant responses for OM than the older two ages, partial $\chi^2(2) = 8.29$, $p = 0.016$. All other interactions of age group and defiant responses were not significant (CC partial $\chi^2(2) = 0$; CM partial $\chi^2(2) = 2.91$, $p = 0.234$; MC partial $\chi^2(2) = 4.33$, $p = 0.12$; CO partial $\chi^2(2) = 2.24$; $p = 0.33$, MM partial $\chi^2(2) = 2.65$, $p = 0.27$; OC partial $\chi^2(2) = 0.50$, $p = 0.78$; MO partial $\chi^2(2) = 0.31$, $p = 0.86$; OO partial $\chi^2(2) = 2.20$, $p = 0.33$). Forgetting the rule would have produced more pronounced effects for younger children and is therefore unlikely. This is further supported by the finding that all ages learnt the rule easily (Section 6.1)

Instead of depending on age, the defiant responses depended on the combination of performance level and rule, with the interactions between rule and defiant responses significant (partial $\chi^2(2) > 6.30$, $p < 0.05$), except for those performance combinations in which defiant responses were rare or absent, see Fig. 4, OO partial $\chi^2(2) = 2.11$, $p = 0.35$, CM partial $\chi^2(2) = 2.91$, $p = 0.23$, and CC partial $\chi^2(2) = 0$. The defiant responses may thus indicate that participants did not always agree with the idea that wins should be rewarded and losses penalized. We return to this point below.

6.4. Reward allocation models

To gain more insight into children's individual strategies, we compared each participant's assigned rewards and penalties to the predictions of different models (Fig. 5).

6.4.1. Model predictions

The **performance model** only considers the target player's individual performance, as can be seen in the upward slope of the single line in Fig. 5. It predicts a penalty when the target player hit the outside ring, an intermediate value for the middle ring, and a reward for the center ring. The model does not take the team win/loss, the different rules, or the other player's performance into account. It is a single factor model without rule differentiation, so does not fit with the group data, but individual participants may still have used this model. From the literature reviewed earlier, rewarding individual performance is within the abilities of even the youngest children tested, and a sizable minority of adults used it in G&L.

The **counterfactual model** (Lewis, 1973) considers whether the target player could have changed the group outcome had he performed differently. This multifactor model considers the rule, the group outcome and both players' contribution to this. When the team lost, yet the target player could have made the team win with a better performance, the model assigns full responsibility for the outcome and a penalty to the target. When the team won, but would have lost if the target player had performed worse, the model similarly assigns full reward. If the outcome is overdetermined, meaning that a change in the target player's performance would not have affected the outcome (due to the other player's performance), the model assigns no responsibility/penalty/reward, predicting a midvalue on the scale. For example, if both players hit the outside ring (OO), the team loses under all conditions. However, the model assigns a penalty only in the additive and disjunctive conditions, since the target player could have made the team win, by hitting the center. In the conjunctive condition, the target could not have made the team win even by hitting the center since *both* players need to at least hit the middle ring and the model predicts no penalty/reward/responsibility. This does not fit with intuition or the

group data, with participants under all rules allocating the largest penalty when both characters perform badly, but again could be used by some individuals.

The **structural model** (Chockler & Halpern, 2004) makes the same predictions as the counterfactual model, except with overdetermined outcomes. As mentioned, for a loss outcome this occurs when both players hit the outside ring (OO) under the conjunctive rule. A win is overdetermined if both players hit the center ring (CC) under the disjunctive rule, since only one player needs to hit the center ring. Instead of predicting zero reward as under the counterfactual model, the structural model predicts some responsibility/reward/penalty, which is more plausible. The model takes the minimum number of changes in the situation, N, for the overdetermined outcomes to become counterfactually dependent on the target and assigns reward as $1/(N+1)$. With two players, one change to the other player (e.g., not hitting the center in the CC disjunctive example) always achieves this, so responsibility/reward/penalty is assigned to each as half. With more players there can be further dilution.

These models were used in G&L's responsibility attribution study and they apply to reward as well. However, although responsibility and reward are closely linked, additional factors may be involved in reward allocation, already suggested by the 'defiant' rewards seen, which the counterfactual/structural models cannot produce. Further models, specifically relevant to reward allocation, were therefore considered.

Given that the players worked as a team, another appropriate reward strategy might be to reward them jointly on that outcome, without regard for individual contributions or responsibility. The **outcome model** thus assigns reward for wins, and penalty for losses. Although determination of the outcome depends both on players' performance and the rule, once it comes to reward allocation itself only the outcome counts. For example, when the target player hit the outside ring and the other player hit the center ring (OC), the model predicts that under the disjunctive rule the target player will be rewarded. Since children had information independent of individual performance to support outcome determination ('duh' or clapping), at a cognitive level, the model may be as simple as the performance model, considering only a single factor for reward allocation, outcome.
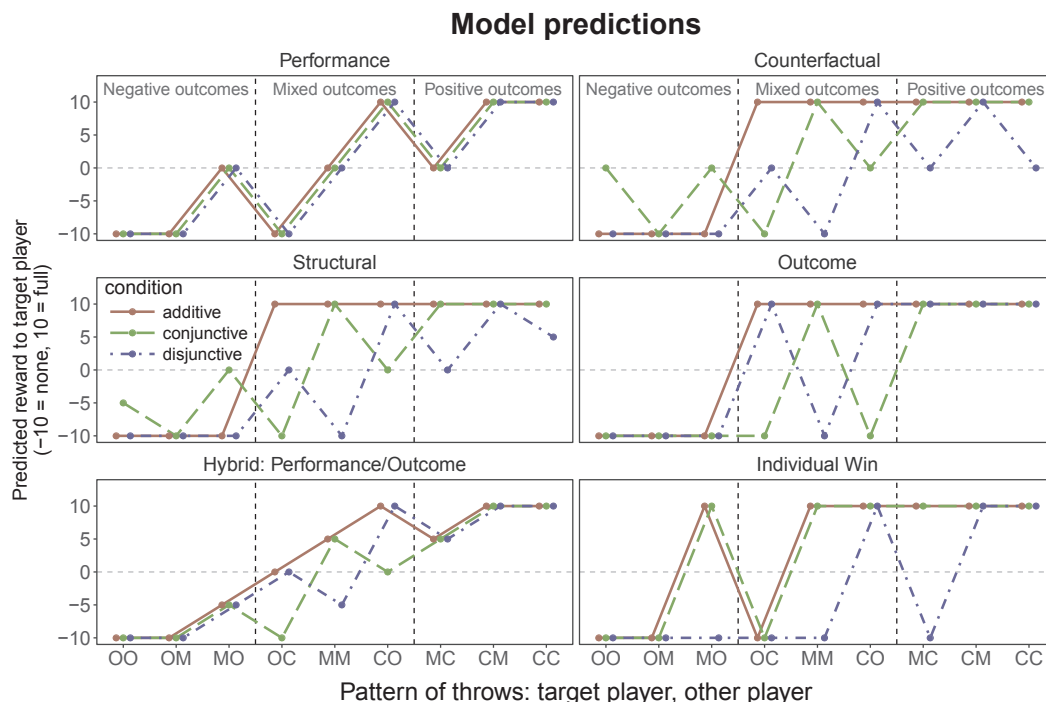
Model predictions

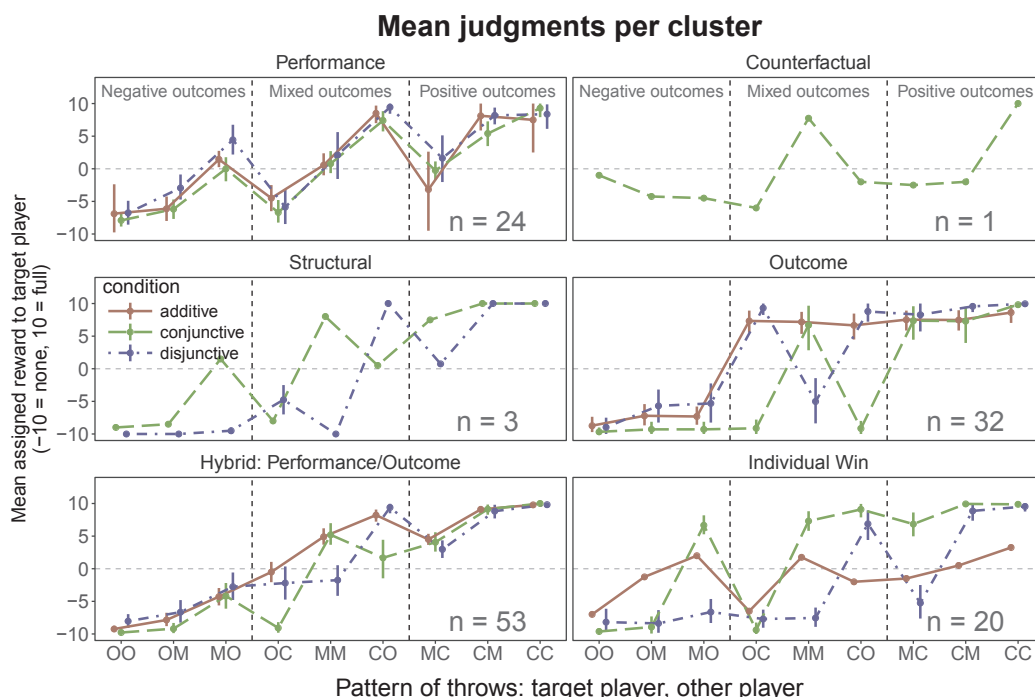**Fig. 5.** Model predictions by six different models.

Fig. 6. Mean judgments for clusters of participants whose judgments were best explained by the respective models. Note: Error bars indicate 95% bootstrapped confidence intervals. n indicates the number of participants whose responses were best explained by each model.

The **hybrid performance/outcome** model, in contrast, considers two factors, the target player's performance and the team outcome, with the response an averaging compromise. For the (OC) case under the disjunctive rule just described, the hybrid model predicts a mid-value, since the team outcome was positive, but the target player performance poor. Fig. 5 shows the simplest case, with equal weights for both dimensions; variants with differential weights are also conceivable. For example, it is possible that children give more weight to the performance factor under one rule or another. Hybrid or averaging models have been successful in many areas of social cognition for adults (Anderson, 1996), and children (Schlottmann, 2000; Schlottmann & Wilkening, 2011; Anderson & Butzin, 1978), and might be a default for multifactor integration when children understand that multiple factors are relevant, but have little knowledge of how the factors fit together.

Finally, the **individual win** model is a simple counterfactual model specific to reward allocation. It considers target player and rule, but ignores the other player's performance and actual outcome, assigning reward by considering what the team outcome would have been if the other player had performed just like the target. It fully rewards the target player when the target fulfills the rule criterion, regardless of non-target performance and actual outcome, otherwise the target is penalized. This allows participants to take the rule into account without penalizing the target player for poor performance of the other player. For example, if the target player hits the middle ring, the model predicts a reward in the additive and conjunctive rule condition, since the team wins if the other player had done the same. In the disjunctive rule condition, it predicts a reward only if the target player hit the center.

When all the models in Fig. 5 are compared to the mean group data in Fig. 3, the outcome and the simple hybrid models, like the performance model, do not predict rule differentiation for games that win or lose under all three rules (right and left panels). Only the structural, counterfactual and individual win models differentiate between rules in all three sectors of the graph, including greater slope for disjunctive on the right and conjunctive on the left, seen in the group data. However, the gradual performance trend across the nine levels of performance does not visually match the more categorical trend in the structural, counterfactual and individual win models. It looks most like the gradual

trend in the hybrid model. Moreover, responses contrary to the group outcome, prominent in Fig. 4, occur only under the performance, hybrid and individual win models. Thus, at a qualitative level, no single model fits the averaged group data fully.

6.4.2. Model fit

To test how well the different models account for individual participants' responses, we regressed each participant's responses across the nine performance patterns with the predictions of the different models. All models except for the hybrid, have two free parameters: one for the intercept in the regression, and one for the weight on the model prediction. The hybrid model has three free parameters: one intercept, one weight for the performance factor, and one weight for the outcome factor. In addition to the six models discussed above, we also considered a **Null model,** which only uses an intercept and thus predicts that participants assign the same value across the different performance patterns. If this model fits best, participants make undifferentiated or unsystematic ratings.[1]

To determine which model best explains each individual participant's responses, we calculated the Bayesian Information Criterion (BIC) score for each of the different models. The BIC takes into account model fit as well as complexity of the model. The hybrid model necessarily achieves a better model fit than the performance or outcome models (since it combines these two). However, the BIC score penalizes for model complexity and a participant will still be assigned to the performance or outcome model as long as the hybrid model's improved fit is small compared to the simpler models. Fig. 6 shows the mean judgments by clusters of participants whose judgments are best explained by the respective models. Table 1 shows how many participants across the different age groups and rules were best explained by each model, with model fits for individuals listed in the appendix.

Note that for the additive rule, the counterfactual, structural, and

---

[1] That a participant's responses fit some model is suggestive but, of course, not conclusive evidence that they actually used this model. It is possible that participants made judgments that are consistent with a model we failed to consider, or that participants choose different strategies over time.

outcome models make the same predictions. Overall, data for 18 participants in the additive condition were equally well explained by these three models, 12 from the youngest age group, five from the 6–7-year-old group and one adult. However, as already mentioned, it is not certain that young children can undertake the complex counterfactual reasoning needed for the counterfactual and structural models, only four participants in total used the structural or counterfactual models in the other two rule conditions and three of these were adults. Moreover, BIC scores for five of these 18 additive participants were 0, indicating that the model perfectly fits the data. Under the other two rules, a BIC score of 0 also occurred for five participants, all in the outcome model cluster, while none of the four participants fitted by the counterfactual/structural model had a 0 score. We assume, therefore, that the 18 participants in the ambiguous additive condition assigned reward based on the win/loss outcome and include them within the outcome model cluster in Fig. 6.

Overall the modeling found large individual differences and no single model was dominant (see Table 1, bottom). The hybrid model (averaging performance and team outcome) explains the largest group of participants' reward assignments (39%, N = 52), followed by the outcome model (24%, N = 32), the performance model (19%, N = 25), and the individual win model (15%, N = 20). Only three participants' responses (2%) were best explained by the structural model, one by the counterfactual and two by the null model.

Importantly, Table 1 shows substantial variations in the model distribution between the three rules: Under the additive rule the hybrid model was dominant (51% of participants), while the performance and individual win model were rare (11% combined). Under the other two rules, the model distribution was more even. The relationship between rule and model fit was significant, $\chi^2(8) = 25.0$, p = 0.002, evidence of rule sensitivity at the individual level.

Age trends, in contrast, were minor. Only under the additive rule was there a change in model fit with age: most of the 4–5-year-olds' data fit the outcome model, but by 6–7 years more than half were explained by the hybrid model, which also explained most of the adult data. This was reflected, for the additive rule participants, in a significant age × rule × model fit association between the outcome model and all other models combined, $\chi^2(2) = 15.8$, p < 0.001. However, there were no clear age differences in model distributions for the conjunctive or disjunctive rule groups. This contrast is further evidence for rule differentiation at the individual level.

Overall, participants' data fit by the single factor performance and outcome models decreased slightly with age, with 54% of 4–5-year-

olds, 44% of 6–7-year-olds, but only 27% of adults' data explained by one of these two models. The more complex multifactor models, especially the hybrid, concomitantly increased with age. However, the association between model fit and age group did not reach significance, $\chi^2(8) = 14.4$, p = 0.071.

It is also worth stressing that the distribution of model fits differed substantially from G&L: Only three of our participants' responses (2%) were best explained by the structural model, whereas the majority of G&L participants fit this model. At the core of the structural model lies the dilution of responsibility in cases of over-determination: Under the disjunctive rule, CC should be rewarded less than CM and under the conjunctive rule OM penalized more than OO. Of 88 participants under the conjunctive or disjunctive rule in our study, only 13 (15%) conformed qualitatively to either of these predictions, in contrast to G&L's findings (cf. Lagnado et al., 2013). This difference between the studies is not due to the child participants here, but appears equally for the adults here: only 1 of 30 adults here diluted responsibility in at least one case of overdetermination.

In sum, modeling individual participant's responses confirmed that children distinguish different causal structures from 4 to 5 years of age. In addition, it suggests there might be differences between reward allocation and responsibility attribution. Reward allocation strategies differed with the rule and with individuals, in contrast to the dominant use of a structural strategy across rules and individuals in responsibility attribution, shown by G&L.

7. General discussion

Two experiments sought to establish how children from age 4 allocate reward and penalty for winning and losing a game to members of a team. Team members' performance varied and games had different rules for winning, so that the same performance could have different effects on the outcome. We found that even the youngest children allocated resources depending on performance, for reward, and for penalty. Most importantly, all ages were sensitive to differences in causal game structures, differentiating their reward/penalty allocations not only when the same performance combination had a different win/loss status, but also when all rules predicted wins (or losses) for the same performance combination. At all ages there were individual differences in reward allocation strategies, with the distribution of strategies depending on the game rule. Our findings suggest potential differences between reward allocation and responsibility attribution. These points are discussed in turn.

Table 1
Percentage of participants across age range and rule whose judgments are best explained by the different models; the modal value in each group is in bold and the number of participants for each group is shown in the end column.

Age	Rule condition	Performance	Individual	Hybrid	Structural	Counterfactual	Outcome	Null model	No. part.
4–5 years	Additive	13%	6%	6%	0%	0%	**75%**	0%	16
	Conjunctive	29%	7%	**50%**	0%	7%	7%	0%	14
	Disjunctive	**33%**	33%	13%	0%	0%	7%	13%	15
	Total	24%	16%	22%	0%	2%	**31%**	4%	45
6–7 years	Additive	13%	0%	**56%**	0%	0%	31%	0%	16
	Conjunctive	**33%**	27%	27%	0%	0%	13%	0%	15
	Disjunctive	21%	21%	**36%**	0%	0%	21%	0%	14
	Total	22%	16%	**40%**	0%	0%	22%	0%	45
Adults	Additive	0%	0%	**93%**	0%	0%	7%	0%	15
	Conjunctive	20%	20%	**33%**	7%	0%	20%	0%	15
	Disjunctive	7%	20%	**33%**	13%	0%	27%	0%	15
	Total	9%	13%	**53%**	7%	0%	18%	0%	45
Combined	Additive	9%	2%	**51%**	0%	0%	38%	0%	47
	Conjunctive	27%	18%	**36%**	2%	2%	14%	0%	44
	Disjunctive	20%	25%	**27%**	5%	0%	18%	5%	44
	Total	19%	15%	**39%**	2%	1%	24%	1%	135

7.1. Reward and penalty

Our study builds on previous research in which children reward based on merit (Anderson & Butzin, 1978; Baumard et al., 2011; Kanngeisser & Warneken, 2012; Sloane et al., 2012; Warneken, Lohse, Melis, & Tomasello, 2011) to more complex contexts, asking participants to consider the team outcome as well as individual performance. Individual character performance was clearly a factor in the reward allocation, and the extra complexity did not lead to difficulties, with participants of all ages rewarding based on merit.

Whether children also take resources away in response to a negative act had not been researched as extensively. Here we found that children also penalize a player for losing a game in proportion to performance. There were no developmental differences.

Further work is needed to determine how allocation of positive and negative reward depends on context. Here, we considered children's allocation of reward and penalty in games, in line with many others (Kenward & Dahl, 2011; Melis et al., 2013), since games, following the work of Piaget (1932) and Kohlberg (1969), are considered safe grounds for children to practice moral and social judgments. Nevertheless, children may treat non-game contexts with potentially more serious consequences in the real world somewhat differently. In particular, future studies should consider whether children react differently when penalizing for a game outcome, or minor social infringement, or when punishing a transgression in which harm is done (Cushman et al., 2013).

7.2. Group structure

Our most important result is that children are sensitive to the differences between causal game structures. This appeared in two ways: First, there were major differences in allocation when the same performance combination won under one rule and lost under another. Children here do not just reward the target character's individual performance, but also the game outcome, with the same individual performance rewarded in a team win, but penalized or rewarded far less in a team loss. Despite initial rule training and on-screen visual and auditory support for the outcome, individual performance levels were the most salient aspect of the experimental trials, so it is a cognitively sophisticated achievement that children from 4 years could interpret individual performance in the context of team outcome.

Second, and even more impressively, the causal structures inherent in the rules even mediated children's allocations when identical performance combinations led to the same winning or losing outcome under all rules; performance was rewarded less when it did not contribute to the win and vice versa. This would seem to reflect an understanding that a given performance could be crucial for a given group outcome under one rule, but less important for the same group outcome under another rule. This subtle concept of the underlying causal structure is evident from 4 years. Ours is the first study to show this, and further research on developmental origins and determinants of this sensitivity to team structure is necessary.

A related implication is that children considered both individual performance and group outcome when applying the merit principle. Ours is one of many studies confirming that children can integrate multiple factors when making graded judgments. Five-year-olds based judgments on two or three dimensions for varied concepts, such as tilt of a beam as a function of weight and distance from the fulcrum (Andrews, Halford, Murphy, & Knox, 2009), volume of a cube as a function of width, height and length (Ebersbach, 2009), or expected utility of a game as a function of probability and one or two possible prizes (Schlottmann & Wilkening, 2011). Two-dimensional judgments appear for children as young as three years (Cuneo, 1980), but of particular relevance here is that 4-year-olds allocate reward considering both merit and need (Anderson & Butzin, 1978; Rizzo & Killen, 2016). Our study similarly shows that distinct merit inputs, from individual

and group, are considered and integrated.

An unexpected result was that children and adults sometimes increased/decreased the initial endowment in contrast to the group outcome. We expected increases in winning games, and decreases in losing games, with no change for players not contributing to this outcome. However, all ages decreased the reward to low scoring players in winning games and increased it for high scoring players in losing games. This type of response, on face value, suggests that children emphasized individual performance and de-emphasized the team outcome.

Forgetting the rule, or not understanding its implications, is not a likely cause of this, since children were trained in the rule, there were perceptual cues, and adults were just as likely as children to do this. Moreover, these responses did not occur indiscriminately, but mainly in cases of discrepancy between individual performance and group outcome. Accordingly, we argue for a different interpretation.

In particular, rather than merely rewarding individual achievement, children and adults may have considered the group outcome while at the same time attempting to redress potential unfairness to an individual, who is being "hard done by" or who "got away with it". This may not always appear in the real world of adults, where the prize is often contingent on winning and the penalty on losing, but with our scale it was possible, and children at least have much experience with consolation prizes given by parents even if there is no official reward. For children this was clearly the more natural approach to responding, and adults, who had no problem adapting to the win/loss constraint of Experiment 1, also switched to this approach in Experiment 2. Under this view, responses contrary to the group outcome highlight that reward allocation is not just a direct response to a character's achievement, but reflects also on goals and beliefs of the rewarder. Remarkably, this active rather than responsive element of reward allocation appears even in a game, and from the youngest age tested. The interpretation is tentative, of course, and needs experimental confirmation.

With the previous point we move from a consideration of group level to individual data. We considered children's underlying strategies and modeled these for individuals. We found pronounced individual differences, with the vast majority at all ages using one of four strategies: rewarding target performance (performance), team outcome (outcome), a hybrid of these two approaches (hybrid), or rewarding as if the outcome only depended on the target player (individual win). Except for the outcome model, all of these can produce the 'defiant' responses discussed above.

In this individual participant modeling, differences between age groups or game rule conditions appear as differences in the distribution of individuals' strategies. As for the group level analyses, we found only minor developmental differences, together with pronounced differences between the different game structures. The latter provide evidence of rule sensitivity at the individual level.

The differences in the distributions of strategies for the three game rules appear meaningful, with participants using more co-operative strategies under more co-operative causal rule conditions. This is clearest for the hybrid strategy, which was by far the most frequently used strategy (52 of 135 participants overall), and which is co-operative, rewarding both players for a win. It was used least often under the least co-operative disjunctive game rule, in which only one player is needed for a win, more under the conjunctive rule, in which both players are needed, and most often in the additive condition, which is most co-operative, due to its compensatory nature. The reverse trend appears for the individual win model, which only focuses on the target player and is most frequent in the disjunctive, least frequent in the additive situation. The other individualistic model (performance) is also least frequent in the additive condition, while the other co-operative model (outcome) is most frequent. The distinction between conjunctive and disjunctive conditions is not clear for the latter two models, however, sample sizes are very small, and the general trend clear. Walzer

(1983) argues that different distributive justice norms are relevant in different societal arenas, and this is reflected here in distinct causal structures creating different contexts that influence the reward allocation strategy.

While the modeling results here suggest that both individual differences and the causal game structure impact how children assign reward, further studies should investigate how individuals adapt to playing under different rules, a question that requires use of a within-subjects design. For example, the hybrid strategy considers both individual performance and team outcome when assigning reward or penalty, and participants might assign different weight to these factors to suit different rules, putting more emphasis on the team outcome for the more co-operative, additive and conjunctive rules, while weighing individual performance more under the more individualistic disjunctive rule. This is a simpler form of rule differentiation than appears at the group level, where allocations recognized the contribution that a particular performance level made to the outcome under a given rule, but could still be used by some individuals.

At a broader level, the cultural context may also influence reward allocate to teams. Studies comparing Asian and Western resource distribution principles often find that equality is preferred over merit-based sharing in Asian and collectivist cultures (Bond, Leung, & Wan, 1982; Rochat et al. 2009; Thomson, 2007). This increases with age, with Singaporean children from five to 17 years moving from equity to equality based distribution (Singh, Chong, Leow, & Tan, 2002). Future studies may also consider whether the integration of individual and group outcome differs in cultures where group harmony has higher social value.

Finally, our results of individual differences in reasoning strategies used for a single causal structure fit with the idea that distributive justice norms are influenced by an individuals' experience (Scott & Bornstein, 2009). It did not seem, however, that individual differences and the influence of individual experience grow with age, rather they were pronounced from a young age. Studies in other domains found close links between children's and parental rules for punishment (Leon, 1984) or between child and parental risk attitude (Levin, Hart, Weller, & Harshman, 2007). Conceivably, reward allocation rules of parents, which presumably are salient to young children, are also transmitted between generations. Future research could investigate the cognitive and social correlates of individuals' reward allocations. These might depend not only on such individual experience, but also on cognitive abilities or on moral orientation towards justice or care (Jaffee & Hyde, 2000).

7.3. Counterfactual reasoning, responsibility and reward

At the inception of our study we had the working hypothesis that reward allocation could serve as a proxy for responsibility attribution, and that it perhaps might be a developmental precursor, internalised with age. We had thus expected that adults' reward allocations would involve the structural strategy found previously for responsibility attribution (G&L), and that children would use this strategy increasingly with age. A developmental trend was expected, since work in other domains suggests that the youngest children here are not capable of the true counterfactual reasoning needed for the structural model.

Studies which suggest children can think counterfactually by the age of four (Harris, German, & Mills, 1996; Riggs, Peterson, Robinson, & Mitchell, 1998) tended to involve a single, short scenario, whereas studies involving multiple possibilities (Beck et al., 2006) or longer, more complex causal chains (German & Nichols, 2003) found that children of that age have difficulty. Indeed, what appears to be counterfactual thinking in these young children could often be basic conditional reasoning (Leahy, Rafetseder, & Perner, 2014; Rafetseder et al., 2010). Rafetseder, Schwitalla, and Perner (2013) discriminated between these types of reasoning in a task with over-determination of an outcome, and < 20% of 5-year-olds gave the correct counterfactual

answer. One would therefore not expect the 4-year-olds here to be capable of the counterfactual thinking necessary for the structural strategy, but we did expect some structural reasoning in the older age group.

Our study clearly showed that these expectations were wrong. Our study was a close parallel to G&L, differing only in content, to make it more child-appropriate (a ball throwing team of two versus a geometric counting team of four), and in the scale used (bipolar reward versus unipolar responsibility scale). However, not even adults used the structural strategy predominant in G&L's responsibility attribution for reward allocation. Counterfactual thinking appeared only in the form of the individual win strategy, a simpler counterfactual that does not require consideration of the consequences of a change in the target's performance. This appeared more often that the expected structural strategy, but still only for a small number (around 15%) of individuals at all ages. More generally, there was no developmental trend towards the adult strategies in reward allocation, rather children behaved for the most part like adults. The different results of the two studies could reflect procedural differences - or conceptual differences between responsibility and reward.

Reward and responsibility are closely linked, of course, such that often the reward for an outcome is based on degree of responsibility for this outcome, but the concepts also differ. For instance, reward has a motivational aspect and is given in part to influence future behaviour, which is less salient in responsibility attribution (Lagnado et al., 2013). The differential emphasis on motivation may, for instance, help explain Cushman et al.'s (2013) finding that when presented with stories involving a character doing harm, either accidentally or intentionally, 4-year-olds deemed the character in both cases responsible for the action, but were less likely to punish accidental harm. Cushman found that children already distinguish, with respect to intent at least, reward and responsibility. Our data fit with this view, but a direct test in our paradigm is still outstanding.

8. Conclusion

This is the first study to consider children's reward allocations in complex causal group structures, with individual performance not necessarily proportional to outcome. It explored how differing causal situations, manifest in differing game rules, influence children's reward attribution. Our study has shown that children's reward allocations are remarkably nuanced from a very early age. Rather than only considering individual merit, children reward the performance in context, taking the team outcome and, even more impressively, the subtle causal differences inherent in the rules into account. Children are social beings, constantly interacting within complex group dynamics both at home and with their peers and it is perhaps not surprising that even at this early age they have learned to navigate complex groups, understanding how different group structures create distinct causal dynamics that necessitate different reaction strategies. Further research will explore how robust these nuanced judgments about distributive justice are, and consider how their origin fits with views including both innate capacity for moral understanding and evaluation (Hamlin, 2013) and a sense of justice that develops with experience (Damon, 1977; Piaget, 1932; Rochat et al., 2009). These initial results demonstrate the complexity of children's moral judgment in group situations, an area of research that has only recently started to be explored.

Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://osf.io/4s6vb/.

Appendix A

See Table A1

Table A1

BIC scores for individual participants in Experiment 2. Lower values indicate better model performance. *Note*: If a model included a negative coefficient, we discarded that model (and set the value to infinite). If a model achieved a perfect fit, we set the value to 0. The 'Best' column indicates which cluster a participant was assigned to. The 'Hybrid' model combines the 'Performance' and 'Outcome' model as described in text and shown in Figs. 5 and 6.

#	Age	Condition	Null	Performance	Counterfactual	Structural	Outcome	Individual	Hybrid	Best
1	4–5 years	additive	53.55	49.07	54.3	54.3	54.3	49.19	Inf	Performance
2	4–5 years	additive	65.38	54.93	65.47	65.47	65.47	63.18	Inf	Performance
3	4–5 years	additive	70.25	68.81	0	0	0	69.88	Inf	Outcome
4	4–5 years	additive	45.39	46.33	44.58	44.58	44.58	45.72	46.72	Outcome
5	4–5 years	additive	50.63	49.1	43.77	43.77	43.77	48.95	45.4	Outcome
6	4–5 years	additive	68.17	67.36	47.39	47.39	47.39	68.7	Inf	Outcome
7	4–5 years	additive	69.81	68.09	52.16	52.16	52.16	70.18	54.06	Outcome
8	4–5 years	additive	68.21	66.51	31.26	31.26	31.26	67.44	33.07	Outcome
9	4–5 years	additive	68.94	67.5	16.01	16.01	16.01	68.71	Inf	Outcome
10	4–5 years	additive	70.32	68.87	0	0	0	69.93	Inf	Outcome
11	4–5 years	additive	62.51	61.53	50.78	50.78	50.78	62.1	52.91	Outcome
12	4–5 years	additive	64.44	63.39	54.11	54.11	54.11	63.8	56.17	Outcome
13	4–5 years	additive	67.05	65.27	37.88	37.88	37.88	66.21	39.7	Outcome
14	4–5 years	additive	65.09	64.68	57.04	57.04	57.04	64.41	59.21	Outcome
15	4–5 years	additive	69.24	60.29	60.52	60.52	60.52	64.71	53.27	Hybrid
16	4–5 years	additive	51.97	49.01	53.85	53.85	53.85	45.94	Inf	Individual
17	4–5 years	conjunctive	63.92	53.39	64.63	63.89	65.83	57.51	Inf	Performance
18	4–5 years	conjunctive	64.64	32.09	60.8	59.86	63.78	57.06	Inf	Performance
19	4–5 years	conjunctive	64.21	51.3	60.77	59.75	62.24	59.68	52.54	Performance
20	4–5 years	conjunctive	59.3	52.85	57.78	56.9	59.58	53.7	55.05	Performance
21	4–5 years	conjunctive	59.76	59.94	56.79	57.17	57.01	60.04	59.01	Counterfactual
22	4–5 years	conjunctive	71.27	70.26	59.68	57.34	0	68.87	Inf	Outcome
23	4–5 years	conjunctive	64.89	59.14	57.52	54.56	55.95	59.13	52.65	Hybrid
24	4–5 years	conjunctive	66.17	57.1	56.75	52.21	55.08	58.4	41.4	Hybrid
25	4–5 years	conjunctive	71.27	63.35	63.32	60.92	64.27	64.64	58.59	Hybrid
26	4–5 years	conjunctive	66.62	60.37	57.1	52.14	53.31	59.48	46.2	Hybrid
27	4–5 years	conjunctive	67.66	41.37	61.44	59.04	65.44	50.25	40.9	Hybrid
28	4–5 years	conjunctive	67.81	65.28	55.93	52.23	38.5	64.07	34.98	Hybrid
29	4–5 years	conjunctive	66.9	56.17	54.47	48.86	56.45	56.56	36.1	Hybrid
30	4–5 years	conjunctive	69.67	58.8	63.34	60.18	67.53	23.31	59.86	Individual
31	4–5 years	disjunctive	62.23	57.09	62.32	62.35	64.38	57.35	Inf	Performance
32	4–5 years	disjunctive	68.44	56.23	67.31	66.55	68.88	63.95	Inf	Performance
33	4–5 years	disjunctive	68.68	57.35	69.35	69.03	70.11	67.85	Inf	Performance
34	4–5 years	disjunctive	63.47	54.94	60.77	59.41	62.34	56.98	56.72	Performance
35	4–5 years	disjunctive	58.18	44.82	54.5	53.12	56.65	50.77	46.47	Performance
36	4–5 years	disjunctive	69.19	68.31	58.68	56.73	41.33	67	Inf	Outcome
37	4–5 years	disjunctive	64.87	54.37	54.71	50.55	57.62	53.27	47.44	Hybrid
38	4–5 years	disjunctive	66.61	62.74	57.67	54.35	50.68	62.46	48.44	Hybrid
39	4–5 years	disjunctive	66.04	48.44	61.16	58.26	63.9	46.54	49.34	Individual
40	4–5 years	disjunctive	65.2	57.57	61.65	59.38	63.93	50.63	59.21	Individual
41	4–5 years	disjunctive	59.08	54.49	52.91	51.91	56.44	51.7	54.86	Individual
42	4–5 years	disjunctive	66.71	61.74	62.73	59.72	63.03	56.45	61.23	Individual
43	4–5 years	disjunctive	70.03	59.77	63.24	59.92	67.63	8.71	60.59	Individual
44	4–5 years	disjunctive	50.55	50.88	51.57	51.49	51.21	52.16	52.67	Null
45	4–5 years	disjunctive	55.65	57.55	57.5	57.39	56.1	57.84	Inf	Null
46	6–7 years	additive	68.22	53.89	68.33	68.33	68.33	64.85	Inf	Performance
47	6–7 years	additive	63.11	42.67	60.43	60.43	60.43	53.66	43.28	Performance
48	6–7 years	additive	68.53	66.68	33.74	33.74	33.74	67.86	35.21	Outcome
49	6–7 years	additive	62.95	60.37	45.56	45.56	45.56	61.68	46.54	Outcome
50	6–7 years	additive	70.17	68.76	0	0	0	69.83	Inf	Outcome
51	6–7 years	additive	70.32	68.87	0	0	0	69.93	Inf	Outcome
52	6–7 years	additive	62.57	63.01	51.94	51.94	51.94	63.71	Inf	Outcome
53	6–7 years	additive	66.83	60.6	48.36	48.36	48.36	62.43	35.89	Hybrid
54	6–7 years	additive	65.89	54.91	55.11	55.11	55.11	60.27	38.25	Hybrid
55	6–7 years	additive	66.34	62.64	46.85	46.85	46.85	65.42	45.32	Hybrid
56	6–7 years	additive	63.47	59.93	46.45	46.45	46.45	61.94	45.88	Hybrid
57	6–7 years	additive	63.3	52.85	55.92	55.92	55.92	57	47.25	Hybrid
58	6–7 years	additive	68.8	55.04	63.54	63.54	63.54	57.12	51.94	Hybrid
59	6–7 years	additive	65.28	45.65	59.88	59.88	59.88	59.88	36.85	Hybrid
60	6–7 years	additive	64.04	41.6	59.47	59.47	59.47	55.58	34.19	Hybrid
61	6–7 years	additive	65.66	56.77	61.04	61.04	61.04	55.84	55.71	Hybrid
62	6–7 years	conjunctive	65.31	55.22	62.25	61.41	65.22	57.33	57.42	Performance
63	6–7 years	conjunctive	58.94	54.19	59.83	58.96	59.33	57.97	56.37	Performance
64	6–7 years	conjunctive	59.82	51.63	59.24	57.6	58.66	55.73	53.36	Performance
65	6–7 years	conjunctive	64.21	46.64	60.87	59.19	62.35	56.9	47.94	Performance
66	6–7 years	conjunctive	66.66	27.75	64.12	62.89	66.2	59.18	Inf	Performance

(*continued on next page*)

Table A1 (*continued*)

#	Age	Condition	Null	Performance	Counterfactual	Structural	Outcome	Individual	Hybrid	Best
67	6–7 years	conjunctive	67.4	65.46	57.93	55.78	43.35	65.19	44.13	Outcome
68	6–7 years	conjunctive	71.27	70.26	59.68	57.34	0	68.87	Inf	Outcome
69	6–7 years	conjunctive	67.04	25.92	62.39	60.56	65.21	56.18	21.16	Hybrid
70	6–7 years	conjunctive	70.14	68.3	57.82	54.97	26.64	67.17	17.9	Hybrid
71	6–7 years	conjunctive	67.03	62.43	56.59	53.92	47.84	63.63	40.48	Hybrid
72	6–7 years	conjunctive	66.15	31.03	62.05	59.97	64.36	54.74	30.37	Hybrid
73	6–7 years	conjunctive	68.86	52.37	63.14	60.6	67.05	47.04	53.75	Individual
74	6–7 years	conjunctive	70.17	59.9	62.97	59.75	67.77	0	60.71	Individual
75	6–7 years	conjunctive	68.81	51.36	62.52	59.66	66.76	40.26	52.4	Individual
76	6–7 years	conjunctive	63.94	53.76	55.25	49.17	56.7	46.08	47.14	Individual
77	6–7 years	disjunctive	67.73	0	64.04	62.49	66.72	57.45	Inf	Performance
78	6–7 years	disjunctive	60.75	47.22	60.67	60.02	61.42	58.58	Inf	Performance
79	6–7 years	disjunctive	63.05	34.48	60.67	59.6	63.19	54.44	Inf	Performance
80	6–7 years	disjunctive	64.4	61.91	59.23	57.24	56.78	61	57.47	Outcome
81	6–7 years	disjunctive	71.27	70.26	59.68	57.34	0	68.87	Inf	Outcome
82	6–7 years	disjunctive	67.53	66.4	58.49	56.96	48.09	65.66	50.12	Outcome
83	6–7 years	disjunctive	65.39	53.98	53.44	48.37	55.18	55.61	30.59	Hybrid
84	6–7 years	disjunctive	69.49	61.47	59.26	55.4	59.52	61.73	51.32	Hybrid
85	6–7 years	disjunctive	64.06	56.76	58.54	56.83	59.67	58.27	55.52	Hybrid
86	6–7 years	disjunctive	65.27	59.25	51.29	45.52	47.83	58.5	33.39	Hybrid
87	6–7 years	disjunctive	66.99	57.58	56.54	50.49	58.47	53.07	48.87	Hybrid
88	6–7 years	disjunctive	65.26	51.34	59.34	56.1	62.08	47.97	50.94	Individual
89	6–7 years	disjunctive	70.06	59.91	62.82	59.64	67.7	4.7	60.76	Individual
90	6–7 years	disjunctive	68.2	58.25	59.88	56.23	63.71	51.39	56.54	Individual
91	Adults	additive	70.32	68.87	0	0	0	69.93	Inf	Outcome
92	Adults	additive	68.07	63.7	43.96	43.96	43.96	65.1	35.4	Hybrid
93	Adults	additive	66.12	58.67	50.78	50.78	50.78	62.52	38.3	Hybrid
94	Adults	additive	68.54	63.33	46.42	46.42	46.42	66.36	33.27	Hybrid
95	Adults	additive	66.86	53.23	59.74	59.74	59.74	59.74	45.41	Hybrid
96	Adults	additive	66.25	59.04	50.46	50.46	50.46	63.56	38.3	Hybrid
97	Adults	additive	67.25	62.9	42.88	42.88	42.88	65.63	34.03	Hybrid
98	Adults	additive	68.06	53.74	62.77	62.77	62.77	55.88	50.34	Hybrid
99	Adults	additive	53.4	44.07	44.45	44.45	44.45	49.95	36.3	Hybrid
100	Adults	additive	67.35	51.82	61.41	61.41	61.41	62.35	45.8	Hybrid
101	Adults	additive	60.33	48.18	56.18	56.18	56.18	52.47	47.37	Hybrid
102	Adults	additive	63.38	54.1	51.09	51.09	51.09	60.13	37.39	Hybrid
103	Adults	additive	65.51	57.2	50.22	50.22	50.22	62.11	30.38	Hybrid
104	Adults	additive	62.44	48.07	55.91	55.91	55.91	52.2	41.31	Hybrid
105	Adults	additive	64.75	51.59	54.66	54.66	54.66	59.17	25.09	Hybrid
106	Adults	conjunctive	59.98	44.98	56.7	54.33	58.55	46.29	46.73	Performance
107	Adults	conjunctive	64.47	32.47	61.38	59.51	62.95	55.57	33.69	Performance
108	Adults	conjunctive	66.51	12.84	63.15	61.77	65.68	57.46	Inf	Performance
109	Adults	conjunctive	66.51	58.65	52.44	42.65	55.02	52.74	45.13	Structural
110	Adults	conjunctive	59.39	57.18	54.49	53.1	49.71	58	50.69	Outcome
111	Adults	conjunctive	69.56	68.42	58.19	55.95	27.17	67.2	29.14	Outcome
112	Adults	conjunctive	71.27	70.26	59.68	57.34	0	68.87	Inf	Outcome
113	Adults	conjunctive	66.86	47.51	57.89	54.17	61.45	52.83	32.55	Hybrid
114	Adults	conjunctive	66.89	54.9	54.28	47.82	57.74	53.48	37.32	Hybrid
115	Adults	conjunctive	64.86	42.3	58.83	56.44	62.13	51.36	41.03	Hybrid
116	Adults	conjunctive	68.31	58.65	56.98	52.87	58.09	60.02	44.49	Hybrid
117	Adults	conjunctive	67.62	63.85	59.41	57.98	52.6	65.63	50.91	Hybrid
118	Adults	conjunctive	68.5	48.18	62.52	59.84	66.57	44.41	49.27	Individual
119	Adults	conjunctive	67.05	46.14	60.75	57.96	64.83	43.92	46.63	Individual
120	Adults	conjunctive	69.41	56.35	61.61	57.97	66.28	35.67	56.15	Individual
121	Adults	disjunctive	58.28	38.03	55.26	53.15	57.58	42.82	40.23	Performance
122	Adults	disjunctive	68.81	60.76	53.62	43.02	57.18	55.05	46.41	Structural
123	Adults	disjunctive	69.36	58.97	57.32	50.99	61.64	52.74	51.01	Structural
124	Adults	disjunctive	57.63	54.73	52.91	50.68	46.78	55.38	47.04	Outcome
125	Adults	disjunctive	70.82	69.71	59.24	56.92	6.12	68.42	7.18	Outcome
126	Adults	disjunctive	71.27	70.26	59.68	57.34	0	68.87	Inf	Outcome
127	Adults	disjunctive	62.43	60.35	53.58	52.1	43.74	60.6	44.67	Outcome
128	Adults	disjunctive	62.99	37.62	58.53	56.96	60.67	55.39	37.05	Hybrid
129	Adults	disjunctive	50.41	45.41	48.24	46.58	46.41	47.81	44.6	Hybrid
130	Adults	disjunctive	66.2	18.08	61.68	59.89	64.5	55.35	4.67	Hybrid
131	Adults	disjunctive	62.91	42.47	56.98	53.97	60.2	44.19	41.85	Hybrid
132	Adults	disjunctive	64.51	39.01	58.39	55.85	61.99	48.42	37.56	Hybrid
133	Adults	disjunctive	70.25	60.2	63.36	60.13	67.93	0	61.08	Individual
134	Adults	disjunctive	68.27	50.15	61.55	58.52	65.85	41.15	50.55	Individual
135	Adults	disjunctive	64.44	50.72	60.48	57	61.42	47.83	50.6	Individual

References

Anderson, N. H. (1996). *A functional theory of cognition.* Mahwah, NJ: Erlbaum.

Anderson, N. H., & Butzin, C. A. (1978). Integration-theory applied to children's judgments of equity. *Developmental Psychology, 14,* 593–606. http://dx.doi.org/10.1037/0012-1649.14.6.593.

Andrews, G., Halford, G. S., Murphy, K., & Knox, K. (2009). Integration of weight and distance information in young children: The role of relational complexity. *Cognitive Development, 24*(1), 49–60. http://dx.doi.org/10.1016/j.cogdev.2008.07.005.

Baumard, N., Mascaro, O., & Chevallier, C. (2011). Preschoolers are able to take merit into account when distributing goods. *Developmental Psychology, 48*(2), 492–498. http://dx.doi.org/10.1037/a0026598.

Beck, S. R., & Riggs, K. J. (2014). Developing thoughts about what might have been. *Child Development Perspectives, 8*(3), 175–179. http://dx.doi.org/10.1111/cdep.12082.

Beck, S. R., Robinson, E. J., Carroll, D. J., & Apperly, I. A. (2006). Children's thinking about counterfactuals and future hypotheticals as possibilities. *Child Development, 77*(2), 413–426. http://dx.doi.org/10.1111/j.1467-8624.2006.00879.x.

Bond, M. H., Leung, K., & Wan, K. C. (1982). How does cultural collectivism operate? The impact of task and maintenance contributions on reward distribution. *Journal of Cross-cultural Psychology, 13*(2), 186–200.

Chockler, H., & Halpern, J. Y. (2004). Responsibility and blame: A structural-model approach. *Journal of Artificial Intelligence Research, 22,* 93–115. http://dx.doi.org/10.1613/jair.1391.

Cuneo, D. O. (1980). A general strategy for quantity judgments: The height + width rule. *Child Development, 51*(1), 299–301. http://dx.doi.org/10.2307/1129629.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108,* 353–380. http://dx.doi.org/10.1016/j.cognition.2008.03.006.

Cushman, F., Sheketoff, R., Wharton, S., & Carey, S. (2013). The development of intent-based moral judgment. *Cognition, 127,* 6–21. http://dx.doi.org/10.1016/j.cognition.2012.11.008.

Damon, W. (1977). *The Social World of the Child.* San Francisco: Jossey-Bass.

Ebersbach, M. (2009). Achieving a new dimension: Children integrate three dimensions stimulus in volume estimations. *Developmental Psychology, 45*(3), 877–883. http://dx.doi.org/10.1037/a0014616.

Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature, 454*(7208), 1079–1083. http://dx.doi.org/10.1038/nature07155.

German, T. P., & Nichols, S. (2003). Children's counterfactual inferences about long and short causal chains. *Developmental Science, 6*(5), 514–523. http://dx.doi.org/10.1111/1467-7687.00309.

Gerstenberg, T., & Lagnado, D. A. (2010). Spreading the blame: The allocation of responsibility amongst multiple agents. *Cognition, 115,* 166–171. http://dx.doi.org/10.1016/j.cognition.2009.12.011.

Hamlin, J. K. (2013). Moral judgment and action in preverbal infants and toddlers: Evidence for an innate moral core. *Current Directions in Psychological Science, 22*(3), 186–193. http://dx.doi.org/10.1177/0963721412470687.

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. http://dx.doi.org/10.1073/pnas.1110306108.

Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition, 61*(3), 233–259. http://dx.doi.org/10.1016/S0010-0277(96)00715-9.

Jaffee, S., & Hyde, J. S. (2000). Gender differences in moral orientation: A meta-analysis. *Psychological Bulletin, 126*(5), 703.

Kanngeisser, P., & Warneken, F. (2012). Young children consider merit when sharing Resources with others. *PLoS ONE, 7*(8), e43979. http://dx.doi.org/10.1371/journal.pone.0043979.

Kenward, B., & Dahl, M. (2011). Preschoolers distribute scarce resources according to the moral valence of recipients' previous actions. *Developmental Psychology, 47*(4), 1054–1064. http://dx.doi.org/10.1037/a0023869.

Kenward, B., & Östh, T. (2015). Five-year-olds punish antisocial adults. *Aggressive Behavior, 41*(5), 413–420. http://dx.doi.org/10.1002/ab.21568.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.). *Handbook of socialization theory and research.* Chicago: Rand McNally.

Koskuba, K. E., & Gerstenberg, T. (2018, March 9). What's fair? How children assign reward to members of teams with differing causal structures. http://dx.doi.org/10.17605/OSF.IO/4S6VB.

Lagnado, D. A., Gerstenberg, T., & Zultan, R. (2013). Causal responsibility and counterfactuals. *Cognitive Science, 47,* 1036–1073. http://dx.doi.org/10.1111/cogs.12054.

Leahy, B., Rafetseder, E., & Perner, J. (2014). Basic conditional reasoning: How children mimic counterfactual reasoning. *Studia Logica, 102*(4), 793–810. http://dx.doi.org/10.1007/s11225-013-9510-7.

Leon, M. (1984). Rules mothers and sons use to integrate intent and damage information in their moral judgments. *Child Development, 55*(6), 2106–2113. http://dx.doi.org/10.2307/1129783.

Levin, I. P., Hart, S. S., Weller, J. A., & Harshman, L. A. (2007). Stability of choices in a risky decision-making task: A 3-year longitudinal study with children and adults. *Journal of Behavioral Decision Making, 20*(3), 241–252. http://dx.doi.org/10.1002/bdm.552.

Lewis, D. (1973). Causation. *Journal of Philosophy, 70*(17), 556–567. http://dx.doi.org/10.2307/2025310.

McAuliffe, K., Jordan, J. J., & Warneken, F. (2015). Costly third-party punishment in young children. *Cognition, 134,* 1–10. http://dx.doi.org/10.1016/j.cognition.2014.08.013.

McGillicuddy-de Lisi, A. V., Watkins, C., & Vinchur, A. J. (1994). The effect of relationship on children's distributive justice reasoning. *Child Development, 65*(6), 1694–1700. http://dx.doi.org/10.2307/1131288.

Melis, A. P., Altrichter, K., & Tomasello, T. (2013). Allocation of resources to collaborators and free-riders in 3-year-olds. *Journal of Experimental Child Psychology, 114,* 364–370. http://dx.doi.org/10.1016/j.jecp.2012.08.006.

Piaget, J. (1932). *The moral judgment of the child.* New York: Free Press.

Rafetseder, E., Cristi-Vargas, R., & Perner, J. (2010). Counterfactual reasoning: Developing a sense of 'nearest possible world'. *Child Development, 81*(1), 376–389. http://dx.doi.org/10.1111/j.1467-8624.2009.01401.x.

Rafetseder, E., Schwitalla, M., & Perner, J. (2013). Counterfactual reasoning: From childhood to adulthood. *Journal of Experimental Child Psychology, 114*(3), 389–404. http://dx.doi.org/10.1016/j.jecp.2012.10.010.

Riggs, K. J., Peterson, D. M., Robinson, E. J., & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development, 13*(1), 73–90. http://dx.doi.org/10.1016/S0885-2014(98)90021-1.

Rizzo, M. T., Elenbaas, L., Cooley, S., & Killen, M. (2016). Children's recognition of fairness and others' welfare in a resource allocation task: Age related changes. *Developmental Psychology, 52*(8), 1307.

Rizzo, M. T., & Killen, M. (2016). Children's understanding of equity in the context of inequality. *British Journal of Developmental Psychology, 34*(4), 569–581.

Rochat, P., Dias, M. D., Liping, G., Broesch, T., Passos-Ferreira, C., Winning, A., & Berg, B. (2009). Fairness in distributive justice by 3-and 5-year-olds across seven cultures. *Journal of Cross-Cultural Psychology, 40*(3), 416–442. http://dx.doi.org/10.1177/0022022109332844.

Schlottmann, A. (2000). Children's judgements of gambles: A disordinal violation of additive utility. *Journal of Behavioral Decision Making, 13,* 77–89.

Schlottmann, A., & Wilkening, F. (2011). Judgment and decision making in young children. In M. K. Dhami, A. Schlottmann, & M. R. Waldmann (Eds.). *Judgement and decision making as a skill: Learning, development, evolution* (pp. 55–84). Cambridge, UK: Cambridge University Press.

Scott, J. T., & Bornstein, B. H. (2009). What's fair in foul weather and fair? Distributive justice across different allocation contexts and goods. *The Journal of Politics, 71*(3), 831–846. http://dx.doi.org/10.1017/S0022381609090744.

Shaw, A., Choshen-Hillel, S., & Caruso, E. M. (2016). The development of inequity aversion. *Psychological Science, 27*(10), 1352–1359.

Sigelman, C. K., & Waitzman, K. A. (1991). The development of distributive justice orientations: Contextual influences on children's resource allocations. *Child Development, 62*(6), 1367–1378. http://dx.doi.org/10.2307/1130812.

Singh, R., Chong, S. S. K., Leow, H. C., & Tan, R. C. H. (2002). Cognitive and social effects in allocation behavior: A new view on loci of developmental differences. *Asian journal of social psychology, 5*(1), 21–47. http://dx.doi.org/10.1111/1467-839X.00092.

Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness? *Psychological Science, 23*(2), 196–204. http://dx.doi.org/10.1177/0956797611422072.

Smith, C. E., Blake, P. R., & Harris, P. L. (2013). I should but I won't: Why young children endorse norms of fair sharing but do not follow them. *PLoS ONE, 8*(3), e59510. http://dx.doi.org/10.1371/journal.pone.0059510.

Steiner, I. D. (1972). *Group process and productivity.* New York: Academic Press.

Thomson, N. R. (2007). Justice in the home: children's and adolescents' perceptions of the fair distribution of household chores. *Journal of Moral Education, 36*(1), 19–36. http://dx.doi.org/10.1080/03057240601185448.

Walzer, M. (1983). *Spheres of justice: A defense of pluralism and equality.* New York: Basic Books.

Warneken, F., Lohse, K., Melis, A. P., & Tomasello, M. (2011). Young children share the spoils after collaboration. *Psychological Science, 22*(2), 267–273. http://dx.doi.org/10.1177/0956797610395392.

Zultan, R., Gerstenberg, T., & Lagnado, D. A. (2012). Finding fault: Counterfactuals and causality in group attributions. *Cognition, 125*(3), 429–440. http://dx.doi.org/10.1016/j.cognition.2012.07.014.