Wesleyan University The Honors College

Biased reasoning:

The influence of group membership on the automaticity of children's assessments of good and bad actions

by

Gwynne Hunter
Class of 2010

A thesis submitted to the
faculty of Wesleyan University
in partial fulfillment of the requirements for the
Degree of Bachelor of Arts
with Departmental Honors in Psychology

Middletown, Connecticut April, 2010

Acknowledgments

There are many people without whom this project would not have been possible. First and foremost, I would like to thank my thesis advisor Anna Shusterman for helping me through this every step of the way. I am fortunate to have conducted research in her lab for the past few years, which gave me the experience and knowledge necessary to devise and carry out this project. Along the way, Anna has been instrumental in assisting me with ideas, analyses, and setbacks. Thank you for telling me, "It will work."

I am grateful for Mariah Schug's invaluable suggestions and help with participant recruitment, and for Kevin Williarty's technical assistance, without which this experiment would have taken far too long to get running. Thank you to all of the actors took the time to be in my videos and pictures.

I greatly appreciate the participation of Snow Elementary School, Macdonough Elementary School and Northwest Children's Center, and all of the parents who brought their children to the lab. I am grateful to the Psychology Department for assisting with finances, and to the Wesleyan Institutional Review Board for approving this study.

I would like to thank the Blue Lab for making my entire research experience so enjoyable. Working alongside students who are so excited about the field of psychology has been a constant source of motivation and encouragement.

Thanks to friends and family for edits, encouragement and everything else.

Table of Contents

Abstract

Previous research has shown that by age 6, children possess detectable implicit intergroup biases. The current experiment sought to determine 1) how much group association is necessary to produce intergroup biases and 2) whether these biases affect how quickly we assess others' actions. We hypothesized that participants would react more quickly to members of their own group performing good actions and members of another group performing bad actions than their own group being bad and the other group being good. In Experiment 1 participants were placed into minimal groups, after which they watched videos of members from their group and another group performing objectively good or bad actions. Reaction times for deciding whether the actions were good or bad were recorded. Though the child results were insignificant, the reaction time trends followed predicted directions. There was a significant interaction between group and action for adults. Experiment 2 amplified the group membership component by including a competitive prime. An explicit measure was added to ensure that group bias was being elicited. Participants from Experiment 2 and in the combined data responded significantly differently to "good" videos as a function of group, indicating that a competitive mindset strengthened implicit biases. Future studies should determine the level of group association necessary to more robustly affect children's implicit assessments of good and bad actions.

Introduction

In my freshman year of high school, I became a Mink. One of the numerous traditions at my school was to place every student into one of three groups: minks, possums or squirrels. We were connected to the groups in a minimal, arbitrary manner: whichever group our senior mentor belonged to, we belonged to. The group labels themselves were meaningless, unless someone possessed a previous affinity for a particular furry creature. Seated in the school auditorium, we were informed that the groups would be competing against each other in trivia contests and races, solely for competition's sake.

Immediately, a transformation occurred. Pitted against each other, friends became foes and bonds formed between complete strangers. The competitions were petty, but to this day I clearly remember which group I belonged to. Why should an arbitrary group membership, which had nothing to offer me nor had any implications in my life, remain with me for years after loyalty ceased to matter? Why do I still remember which of my good friends were with me, and which were with the "others"? Did those meaningless categorizations affect my perceptions of students in other groups, without me even noticing?

Group bias

Every day, we judge other people's actions. These judgments may be conscious or nonconscious, imposed on clear or ambiguous actions. The individuals whom we judge may come from within or outside of our social groups (Ashburn-Nardo, Voils, & Monteith, 2001; Dunham, Baron, & Banaji, 2008; Tajfel & Turner, 1986).

Currently, one question that is not fully answered by research is the level of group association needed to produce intergroup bias, and the level of bias necessary to affect our judgments. Whether we are aware of their influence or not, do our social groups dictate how quickly we process and judge the actions of others? If so, how personally attached to the group must we be for group membership to affect our judgments? The current study aims to investigate these questions by assessing how long it takes 5-10 year-old children, who have been placed in arbitrary minimal groups, to judge objectively good and bad actions performed by own- and other-group members.

Implicit versus explicit bias

Biases can manifest themselves in two ways: implicitly, outside of our conscious control, and explicitly, within our conscious control. Implicit and explicit biases follow developmentally distinct courses (Dunham, et al., 2008). Explicit preferences for one's own group, or "ingroup," follow a pattern from childhood to adulthood of rising, leveling off, and then decreasing somewhat. White North American children express explicit negative attitudes toward real-life outgroup individuals as early as age three (Baron & Banaji, 2006). By seven years old, the negativity of these explicit attitudes starts to decline, until it reaches adult-like levels at age twelve. Minority group members show the same developmental pattern but with comparatively less ingroup bias, as the natural inclination to favor the ingroup is counteracted somewhat by the influence of majority social groups (Dunham, et al., 2008).

Possibly a facilitator of the eventual decline in their explicit ingroup preference is children's growing awareness of how their opinions compare to others' (Baron & Banaji, 2006). Because of this awareness of others' opinions in relation to theirs, and the desire to mitigate any discrepancy between the two, older children (7-8 year-olds) develop an aversion to inequality. In allocation studies, this translates to them sharing in a more egalitarian manner with their own group and other groups than do younger children (Fehr, Bernhard, & Rockenbach, 2008).

Implicit biases, on the other hand, form later than explicit biases but change little over their developmental trajectory. Implicit attitudes are assessed using Implicit Association Tests (IATs), which measure the associative strength between pairs of concepts, such as social categories and attributes (Dunham, et al., 2008). IATs reveal that implicit biases toward social groups remain relatively constant from the time that children first form detectable attitudes (by age 6) through adulthood (Baron & Banaji, 2006; Dunham, et al., 2008). The different developmental trajectories of implicit and explicit biases may be due to the fact that children can eventually recognize and monitor their explicit biases, but are never able to consciously control their implicit biases (Baron & Banaji, 2006).

Attitudes toward real-life social groups

How and when group bias forms depends fundamentally on how groups are defined. Several studies indicate that children and adults may conceptualize and understand group membership differently. Contributing to this discrepancy is the finding that adults and children consider different characteristics of a person to be more representative of that person's self (Diesendruck & haLevi, 2006).

Adults believe that personality traits, rather than ethnicity, are a better ground for inference about a person. Children consider more superficial status-related qualities, like social status and ethnicity (even in the absence of any physical similarities between supposed group members) to be reliable indicators of a person's character. Children are likely unaware of why seemingly unrelated people are categorized under the same label. Lacking an understanding of social divisions, they must trust that there are good reasons for these categorizations. Children therefore assume that there is some essential, if invisible, quality about a person that warrants assignment to seemingly arbitrary categories (Kinzler, Shutts, & Correll, in press).

Regardless of how they conceptualize social categorization, both children and adults display implicit negative and positive associations with salient social groups (Baron & Banaji, 2006; Dunham, et al., 2008; Tajfel, Billig, Bundy, & Flament, 1971). These biases do not always align with explicit attitudes. For example, most adult participants taking IATs do not consider themselves to be racist, nor do they believe that they judge other races differently in everyday activities (Baron & Banaji, 2006; Glaser & Knowles, 2008). Yet IATs frequently contradict this assumption, revealing that many individuals assess words more quickly or slowly depending on which social groups the words are paired with.

Whites, for example, are often faster at deciding whether they saw the word "Black" when it was paired with "bad" than when it appeared with "good," despite a lack of any explicit racism (Glaser & Knowles, 2008).

IATs are useful for revealing where implicit group biases already exist as well as the relative strengths of those biases. These findings can then be compared to individuals' explicit attitudes toward the same groups. But while existing group associations can be tapped into, they cannot be manipulated. This leaves open the question of how group membership factors into bias development. Also, because of their design, IATs cannot tease apart whether one's own group is viewed positively, the outside group viewed negatively, or both. All that can be determined is that a difference in reaction times exists between assessments of the two groups. To figure out how group bias forms in the first place, and to determine the direction of the bias, group membership must be manipulated in a controlled setting.

Attitudes toward minimal groups

The Minimal Group Paradigm (MGP) is one way to manipulate group membership. In the MPG, participants are assigned to an arbitrary, made-up group void of any connection to real-life social groups. This ingroup is countered by the existence of a second arbitrary categorization, the outgroup, to which the participant does not belong.

Some adult studies suggest that simply being assigned to a group is sufficient to activate positive attitudes toward the ingroup and negative or neutral attitudes toward the outgroup (Ashburn-Nardo, et al., 2001; Otten &

Wentura, 1999; Tajfel, et al., 1971). Assignment to a category may implicitly tap into the propensity to derive self-esteem from the groups to which we belong, leading to an immediate proclivity for bias (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987).

For children, however, the development of bias may depend on what characteristics the minimal social categorizations are based on. A recent MGP study found that visually salient information alone does not appear to promote full-fledged generalization until some time after age six (Baron, Dunham, Banaji, & Carey, under review). Noun labels, on the other hand, are sufficient for formulating distinctions between arbitrary groups. Baron and colleagues (under review) found that having an authority figure provide verbal distinctions between groups was enough to make 3-4 year-old children assume that the groups were essentially different. Once group distinctions have been established, children robustly generalize individual behaviors to other members of those groups (Baron, et al., under review; Bigler, Spears Brown, & Markell, 2001).

Another basis for not only differentiation, but also discrimination, between groups is an authority figure's partial, functional use of group labels (Bigler, Jones, & Lobliner, 1997; Bigler, et al., 2001). In 1968, elementary school teacher Jane Elliott created a minimal group situation using naturally occurring groups: she divided her class based on whether children had brown or blue eyes (for a review, see Stewart, Doan, Gingrich, & Smith, 1998). On some days, Elliott told the class that brown-eyed people were superior to blue-eyed people, and were allowed certain privileges that blue-eyed people were denied. On other days, she reversed the group statuses. By simply telling children "facts" about their groups with no evidence to back up the claims, Elliott was able to elicit strong group discrimination where there had previously been none.

While Elliott's study involved explicit authoritative assessments of group status, intergroup discrimination can arise even when the authority's functional use of group labels is preferentially neutral. In one pair of studies, children at summer school were placed into either a blue or yellow t-shirt group (Bigler, et al., 1997; Bigler, et al., 2001). Some classrooms contained posters implicitly implying that one group was of a higher status than the other. In other classrooms, teachers made non-competitive functional use of the groups, such as having children complete tasks in groups with the same t-shirt color. A third set of classrooms combined the posters and functional use of groups by teachers. Researchers found that functional use of social category labels by teachers, but not biased posters, was enough to produce intergroup discrimination.

A competitive atmosphere, without the unifier of physical similarity or any influence from authority figures, is another means through which to produce substantial intergroup bias (Sherif, Harvey, White, Hood, & Sherif, 1954/1961). Sherif and colleagues' (1954/1961) classic experiment found that when pride and position are at stake, affiliation with one's group intensifies. This finding demonstrates that there need not be any "essential" qualities that members of a group share to encourage group affiliation; simply belonging to a group that seeks to differentiate itself from another faction is enough to cohere the members of a group.

The importance of relative hierarchical group positions is apparent not only in competitive situations but also in MGP allocation experiments. When choosing how to distribute resources to ingroup and outgroup members, participants tend to favor allocating more to their own group than to another group, even when these groups are not biologically relevant (i.e., related or important genetically) and the participants' own future membership in the group is irrelevant or uncertain (Gaertner & Insko, 2000; Grieve & Hogg, 1999; Hertel & Kerr, 2001; Lemyre & Smith, 1985; Nosek, Greenwald, & Banaji, 2007; Tajfel, et al., 1971). What is striking is that this bias persists even in circumstances where allocating more to the outgroup would result in more overall gain for the ingroup (Tajfel, et al., 1971). The ingroup's relative advantage over the outgroup, it seems, takes precedence over an absolute gain for the ingroup.

Though the studies discussed thus far give the impression that group bias takes little impetus to develop, some studies contend that even within real-life social groups, bias is not inevitable. Certain social statuses or categorizations result in lower levels of bias (Ashburn-Nardo, et al., 2001). In both real-life and artificial groups, minority or low-status group membership may compete with the desire to belong to the dominant or high-status social group(s), lessening or eliminating a preference for one group over the other (Baron & Banaji, 2006; Bigler, et al., 2001; Dunham, et al., 2008).

The findings from MGP and IAT research have provided us with valuable information about the existence, development, and expression of group biases

(Baron & Banaji, 2006; Bigler, et al., 1997; Bigler, et al., 2001; Fehr, et al., 2008). The unanswered questions are why this bias forms in the first place, and why it should affect our reasoning to the degree where we form implicit preferences for artificial, biologically irrelevant, meaningless groups. What is the mechanism operating behind bias formation and its incorporation into our reasoning? Research on this topic is in the early stages, but much evidence points to a particular mechanism that may be operating behind our tendencies to categorize and evaluate based on little input: Bayesian reasoning.

Bayesian reasoning

In our everyday lives, we must make decisions, figure out whom to trust, rely on information from others, and formulate beliefs based on limited statistical input. We certainly do not wait until we have witnessed a viable number of instances before making these calculations. If we did, we would be waiting an impossibly long time trying to accumulate enough observations to make a statistically rational assessment of any given situation (Gallistel & King, 2009). For example, if we saw someone stealing clothes from a store, we would not try to observe this person entering the store many times to assess whether their act of stealing was an unusual or common incident; this type of statistically "correct" observation would not even be feasible here. We would quickly make the assumption that this person is familiar with stealing, and make judgments that we personally associate with thieves.

Just because this type of rapid, biased reasoning may not be statistically accurate does not mean it is wrong or impractical. At the core of Bayesian reasoning is the concept of drawing hypotheses and conclusions based on a statistically limited or inoperable amount of input (Gallistel & King, 2009). In the shopping thief case, perhaps we later see the thief sitting next to us in a park. If we are still deciding what to think about that person, we may be less careful about where we leave our valuables. Our biases, though statistically unconfirmed, lead us to err on the side of caution; they form what could be called "common sense." The current study aims to assess our propensity to reason in a Bayesian manner by determining how biases produced by minimal group membership affect our interpretations of others' actions.

Even infants seem capable of understanding that biases can alter the statistical probabilities of expected outcomes (Xu & Garcia, 2008). In Xu and Garcia's study (2008), 8-month-old infants watched an experimenter select a few balls from a larger sample. The selected sample either matched the color of the larger sample (expected outcome) or did not (unexpected outcome). In a random situation, conveyed by the experimenter's averted gaze during selection, infants looked longer at the unexpected outcome. Critically, if the experimenter appeared to be selecting particular balls, infants did not look longer at the "unexpected" outcome, observing no incongruence between the experimenter's desires and the resultant outcome. It therefore seems that from a very young age, children (implicitly) understand that the presence or absence of biases can produce divergent outcomes for the same set of inputs.

A few years after infants recognize the effect of biases on expected outcomes, bias enters into their own reasoning. Evidence for a Bayesian reasoning mechanism has been found in 3-4 year-olds (Sobel, Tenenbaum, & Gopnik, 2004). In Sobel and colleagues' (2004) study, children observed various objects being placed on a plate-like device called the "blicket detector." If an object activated the device, the object was a "blicket." The critical manipulation in the experiment was when two objects, A and B, were placed on the blicket detector. If two objects were placed on the detector and the device did not activate (did not "detect" a blicket), then it was obvious that neither object was a blicket. If the device did activate, the experimenter would place object A on the detector by itself, which would activate the device; B was not placed on the detector by itself.

Importantly, after watching A activate the detector children were asked which object(s) was a blicket, to which they responded that only A was a blicket. Of course, they had no idea whether B was a blicket or not. However, children reasoned that if A+B = activation, and A = activation, there was no reason for B = activation. In other words, they only reasoned as far as was necessary for the observed output to result. Though object B had as much chance as Object A as being a blicket, that deduction was unnecessary and therefore not carried out. This exemplifies how identical input analyzed through Bayesian reasoning, which emphasizes hypothesis testing, and through traditional statistical reasoning can result in different outcomes.

Implications for Bayesian reasoning

As previously stated, if in real life we were to consider the possible outcome of every situation, we would be overwhelmed. Sobel, Tenenbaum, and

Gopnik's (2004) study demonstrated that to deal with this limitless input, we may sometimes reason only as far as is necessary for our observations to align with the outcome; only enough input to produce the resultant perception is considered. The implications for this type of reasoning in real life are considerable.

The following is a hypothetical translation of the Sobel study into a real social situation: Suppose that a child observes two other children (A and B) together stealing a cookie jar. Child A is a member of the observer's social group, and child B is not a member of the observer's social group. The observing child already possesses biases that favor his ingroup. When he sees the two children steal the cookie jar, he knows that it is a bad action. In his mind, however, the child can attribute the bad action to the outgroup member – perhaps it was his idea. There is no reason for him to conclude that just because both children stole the cookie jar (A+B= activation), that each child is a thief. The outgroup member could solely carry the blame (B= activation), leaving the role of the ingroup member ambiguous. Knowing whether this type of reasoning indeed occurs in real-life social situations is essential for future bias reduction efforts.

Present study

The current study combines aspects of IAT measures and minimal group situations in order to address questions that neither line of research has fully answered: How much group induction/association is necessary to produce explicit biases? And do these biases affect our nonconscious assessments of others' actions? The study will simultaneously allow us to examine how

Bayesian reasoning factors into our judgments. Child participants (5-10 year-olds) are near or older than the youngest age documented as able to express differential implicit and explicit biases, but younger than the age at which adult-like monitoring of explicit biases is typically reached (Baron & Banaji, 2006). Adults are included as both a control group (assumedly responding with a full understanding of the task and with mature rapid processing and response capabilities), and as a comparison to children.

Implicit and explicit biases have thus far been measured in a limited number of ways. IATs gauge the level of bias present in relation to real, salient social groups, but they generally focus on existing groups (e.g., nationality, race, gender). Also, IATs can only tell the degree to which participants respond differently to the two groups; they cannot differentiate between an ingroup bias and an outgroup discrimination effect. IAT research is therefore unable to answer the following questions: 1) How do group interactions and memberships produce biases 2) What is the nature of these biases, and 3) How do these biases factor into our reasoning outside of simple associations?

Studies utilizing the Minimal Group Paradigm attempt to answer some or all of these questions. By manipulating group membership with new, arbitrary social categories free of former biases or associations, MGP studies can attempt to determine the level of group association at which these biases form (Gaertner & Insko, 2000; Grieve & Hogg, 1999; Hertel & Kerr, 2001; Lemyre & Smith, 1985; Nosek, et al., 2007). These biases can then be tested both implicitly and explicitly.

MGP allocation studies can establish the level of minimal group association necessary to produce explicit biases, but these studies do not tell us how this bias correlates with implicit attitudes or affects rapid judgments. Some MGP studies incorporate IATs to measure the implicit associations formed by minimal social groups (Ashburn-Nardo, et al., 2001; Baron & Banaji, 2006; Otten & Wentura, 1999; Ray, Way, & Hamilton, 2009). However, these studies only reveal the associations resulting from assignment to a group. They do not reveal further reasoning or information processing on the basis of group membership. The current study aims to investigate this question by combining minimal groups with a timed assessment of others' actions, an innovation in the implicit association and minimal group literature.

This experiment creates a minimal group situation by placing participants in one of two groups. They watch videos of own- and other-group actors who perform objectively good or bad actions to a neutral second party. Participants must quickly assess whether the actions are good or bad, without time to consciously consider which group they belong to in relation to which group they see. Afterward (in Experiment 2), an explicit measure assesses conscious bias toward the two groups.

The important measure is how long it takes children and adults to make these decisions. This will reveal how much association with minimal groups is sufficient to produce judgment-altering bias. Imagine what the implications would be if simply being placed in an arbitrary, biologically irrelevant group with only the most minimal group induction led to a difference in the time it took children to judge obviously good or bad actions. What would that say about how children judge others' actions in real life, in the context of real social groups and potentially much higher levels of bias?

Besides using reaction times to measure assessments of actions in an MGP, another innovation of this experiment is that there is no interaction between groups; instead, each actor is conceptualized in relation to the participant rather than in relation to other actors. In an IAT, interactions between the ingroup, outgroup, good and bad categories are analyzed simultaneously. This makes it impossible to tease apart reactions to ingroup good, ingroup bad, outgroup good and outgroup bad categories. Instead, the relative reactions to those four components are masked behind a generalized division between reactions to the two social groups. In the present experiment, only one group and one action type are presented at the same time. Any effects are therefore due to an interaction between the participant's group and the group on the screen. This allows the four conditions of group by action to be analyzed in comparison to each other. Separating the conditions makes it possible to tease apart the relative strengths of each component, and assess whether group bias is more a result of ingroup favoritism or outgroup discrimination.

A final innovative factor is that this experiment combines realistic situations with unknown groups, using real actors as members of the artificial groups. In IATs, the social groups are real, but there is little realistic interface in the black-and-white still photographs used in the tasks (Goldinger, He, & Papesh, 2009). By having participants wear t-shirts and wristbands, this experiment involves participants directly, increasing the chance that they will find the situation realistic and related to them.

Hypotheses

This experiment could produce a number of different results. This study looks at how biases factor into reasoning outside of simple associations, and the nature of this bias (favoritism or discrimination). Previous MGP research indicates that simply being placed in an arbitrary group can create ingroup favoritism in adults. In this experiment, that could translate to faster reaction times to all ingroup actions, due to higher attention to and faster processing of ingroup members (Baron & Banaji, 2006; Glaser & Knowles, 2008; Nosek, et al., 2007).

Within the reactions to ingroup members, it is likely that participants will have faster reaction times for ingroup good actions but slower reactions for ingroup bad actions, due to the desire to characterize one's own group as good and conversely, a reluctance to characterize one's own group as bad. A difference in reaction times between ingroup and outgroup actions would indicate Bayesian reasoning; the actors' identical actions, and participants' lack of experiences with members of either group, mean that theoretically there should be no reason to respond to one group's actions differently. Any difference would therefore be due to group membership biases.

The experiment could also produce no significant differences in reaction times. Due to disparate experimental designs in the child literature, it is unknown exactly how much group induction is needed for the development of detectable implicit biases; determining this level is another goal of the study. It is also uncertain whether action assessments capture bias differently than do association measures. This experiment starts with very basic group membership and increases salience of groups as needed. The ability to manipulate group induction will allow us to determine how much group association is necessary for biases to arise, as well as to examine the implicit and explicit nature of these biases.

Experiment 1

Method

Participants

There were 22 child participants (12 males; mean age 6;11) and 19 adult participants (7 males; mean age 21;1) tested. No child participant was removed from the analyses. One adult participant's data were discarded due to disregard for experimental instructions. Child participants were recruited through a database of local families, as well as through local kindergartens and elementary schools. Testing occurred in the Wesleyan Cognitive Development Laboratory and at local schools. Adult participants were students at Wesleyan University.

Stimuli

Training videos. Stimuli consisted of videos of young adults (Wesleyan University students) engaged in "good" or "bad" tasks. Children viewed eight videos to become accustomed to pressing the keys for the task. In the practice videos, both actors wore white t-shirts.

There were two story frames: candy and drawing. In the candy videos, one actor either shared candy with or stole candy from the other actor. In the drawing videos, one actor either complimented or scribbled on the other actor's drawing (see Appendix 3 for more detailed description of the videos). Actors and sides of the screen were counterbalanced within subject. Videos were enhanced to be brighter. Good and bad candy videos were edited to be equal in length, as were good and bad drawing videos.

Test videos. Stimuli consisted of videos of young adults (Wesleyan University students) engaged in "good" or "bad" tasks for two story frames: puzzle and fall. The puzzle videos involved a neutral (white shirt) person putting a puzzle together. The target person (wearing a red or blue t-shirt) either helped the actor complete the puzzle or destroyed the puzzle. In the fall videos, a neutral person fell while walking. The target person either laughed at the fallen person or lent a hand (see Appendix 3 for more description of the videos).

One actress played the neutral person in all of the puzzle videos, and another played the neutral person in all of the fall videos. Each target person was filmed twice, once in each color t-shirt, to control for effects of person. This created 16 videos, 8 of which were shown to each participant (with no repeated actors). All falling videos were edited to the same length (6.6 seconds) and all puzzle videos were edited to the same length (10.0 seconds). Videos were all enhanced equally to minimize effects of backlighting and to bring out the t-shirt colors.

Photographs. Stimuli consisted of four chest-to-head photographs: two females (one blue and one red) and two males (one blue and one red). To control for any effect of person, individuals in the photographs were counterbalanced across participants so that some participants saw particular individuals wearing blue shirts and some saw them wearing red.

Materials

Keyboard cover. A manila folder was cut to create two keyboard covers, one with the smiley face on the left and the frown on the right and one with the reverse set-up. Square cutouts were created above the D and K keys, and a line drawing smile or frown was drawn at a set height above each key. The covers were taped to both sides of the laptop to prevent movement during testing.

Stimulus presentation*.* A Macbook laptop was used to present stimuli, with an external keyboard attached to the right side of the device for the experimenter's use (so that the keyboard cover was not moved during testing). PsyScope X B53, free Psychology experimental software, was used to play the videos and capture reaction times.

Procedure

Before the trial videos, children were told that they would see videos of people doing good or bad things. They were instructed on which keys to press, how often, and that they could press the key as soon as they thought they knew whether the action was good or bad (see Appendix 1 for full instructions).

After these instructions, comprehension checks were given. The children were then shown the practice videos. Most children understood the task, but for those who did not perform correctly (either by pressing the keys too many times or giving incorrect responses), feedback was given during this period.

Following the practice period, the computer was moved to one side of the testing table, and the experimenter assigned the participant to either the red or blue group. The child was given a red or blue t-shirt and matching wristbands to emphasize attachment to the group to the actors in the videos. Wristbands were given to provide a visual cue of the child's color group that they could see at all times, and to ensure that children who did not want to wear the t-shirt would still feel connected to the group (six participants chose to hold the t-shirts in their laps or across their chests instead of wearing them). No explanation or further description of the groups was given, and children for the most part did not comment on group membership.

The photographs were introduced after the participants put on their t-shirt and wristbands. The photographs served as measures of comprehension and means of group reinforcement, as well as an introduction to conceptualizing the groups as "your" and "other" in addition to "red" and "blue."

After the photographs were shown, the computer was brought back in front of the child. The instructions for which keys to press and how often to press them were repeated. No mention was made of the groups. Children were then shown one video set consisting of eight test videos. There were four video sets, counterbalanced across participants: one random order for each set of 8 videos, and the reverse of each of these orders.

Using normalized scores in analyses

Using normed or raw data makes no difference when comparing reaction times within videos, as normalized scores are proportional to raw scores and therefore produce the same statistical results. With perfect counterbalancing, this is also not a problem when comparing across videos.

However, due to the nature of participant recruitment, we could not perfectly balance participant sex, video set and group (16 combinations). Also, there was a possibility that participants' response times would change across trials. For example people might respond more quickly as they got more familiar with the task. Because it was not possible to fully counterbalance for order along with all other dimensions, we had to account for the possibility of order effects in our analyses.

Assume for now that first videos in a sequence are always reacted to the most slowly due to unfamiliarity with the videos and/or task. Imagine averaging raw reaction times to a first and a middle video to compute the ingroup good RT for one participant. Compare this to another participant's ingroup good RT, which requires averaging two middle videos. The first pair would produce a higher average reaction time not due to group but instead due to video order.

Norming the scores by converting raw reaction times to z-scores keeps the proportions within videos intact; that is, it makes no difference when comparing first videos to each other whether they are in z-scores or raw scores, because the relative differences between the videos are the same. Normed scores, however, allow for comparison across videos (unlike raw scores, as illustrated in the previous paragraph). In other words, norming translates the first and middle videos to the same scale, which means that when they are averaged, they can be reliably compared to the average of the two middle videos. Any differences that arise will be due to a factor like group rather than an unrelated factor like video order.

Z-socres were computed for each specific video (e.g., "blue good female puzzle"). The z-scores were averaged for each participant into participants' ingroup good, ingroup bad, outgroup good and outgroup bad (two videos for each condition).

Because scores were normed for a particular set of raw data, each different grouping (Experiment 1, Experiment 2, Adult, etc.) was renormed to accurately reflect the raw data contained in that data set.

Analyses

Repeated Measures ANOVAs were conducted to test for effect of Group (In, Out), Action (Good, Bad), or interaction between group and action (Group*Action). Mean z-scores for were computed for ingroup good (OG), ingroup bad (IB), outgroup good (OG) and outgroup bad (OB) within data sets to see if any trends were present. Paired-samples t-tests were used to test for differences between IG, IB, OG and OB video pairs. An alpha level of .05 was used as a measure of significance in all analyses.

Because previous implicit experiments had only tested ages 6 and older, we were unsure whether 5-year-olds would exhibit expected task comprehension and performance. Analyses were conducted with 5-year-olds, without 5-year-olds, and with only 5-year-olds to test for any difference in responses based on age. The results did not differ when analyzed with or without 5-year-olds, nor did 5-year-olds differ as a group. Age differences on the implicit measure will therefore not be discussed any further.

Accuracy

Because of participants' high level of accuracy when responding to the videos, all responses were included in the analyses.

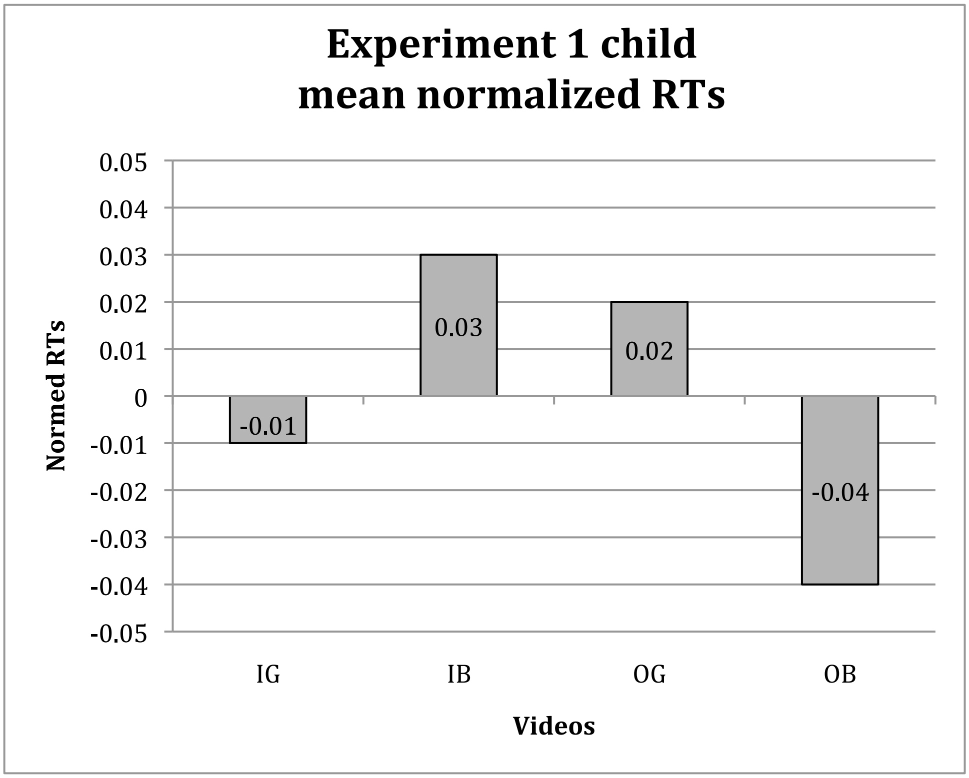
Between Experiment 1 and 2, 4% of trials (out of 336 total) were discarded due to a technical malfunction where a participant pressed a response key twice during a video, and the second key press was recorded for the next video (overwriting any presses intended for the next video). These data points were infrequent and were easy to detect amongst the other data and delete.

Results: Children

Trends in reaction times

First, we wanted to see whether the mean reaction times (RTs) in Experiment 1 aligned with our predictions, namely that ingroup good and outgroup bad would be consistent with biases and therefore produce faster RTs, while ingroup good and outgroup good would be inconsistent with biases and produce slower RTs. Figure 1 presents the average z-scores across participants for IG, IB, OG and OB videos (negative numbers represent faster RTs). Thus, the trends followed the predicted directions for RTs.

Figure 1. Mean z-scores for Experiment 1 child reaction times. For purposes of clarity, standard error bars are not included in the figures. SE values for Experiment 1 were 0.18 (IG, OG, and OB) and 0.20 (IB).



Effects of group and action

We first analyzed whether there was a statistically significant difference in RTs as a function of group or action. A Repeated Measures ANOVA revealed no significant effects of group (In, Out; $F(1, 21) = 0.11$, $p = .74$), action (Good, Bad; $F(1, 21) < 0.00$, $p = .99$), or interaction between group and action (InOut * GoodBad; $F(1, 21) = 0.57$, $p = .45$).

Differences between video pairs

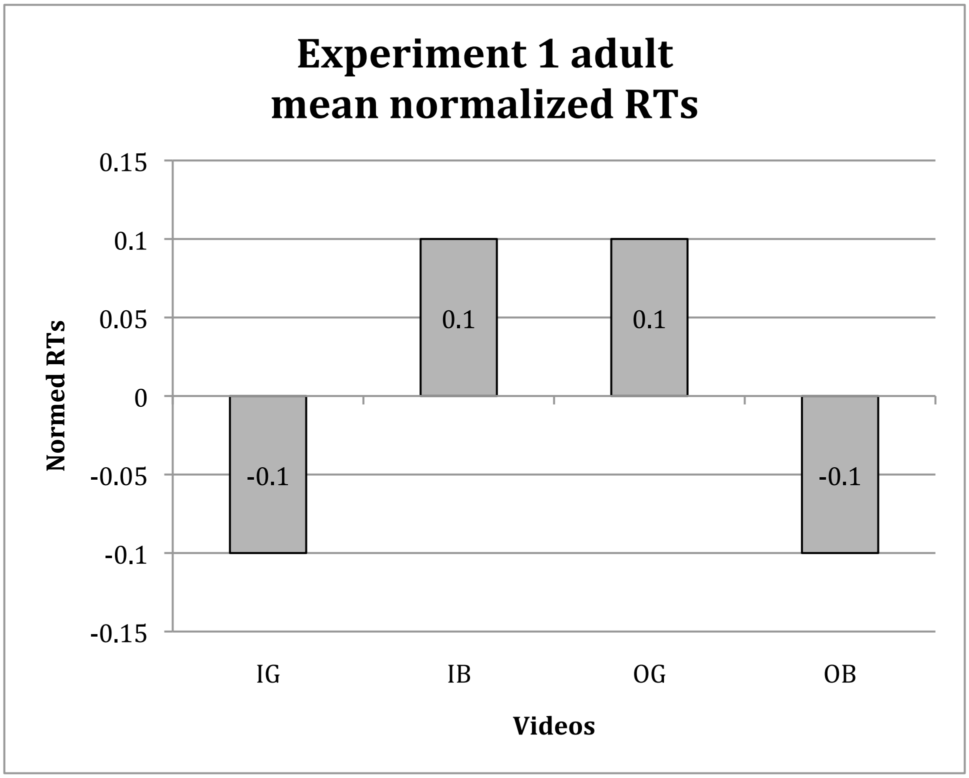
Because the ANOVAs likely were limited in their power to detect an effect given the overall variance in the data, paired-samples t-tests were conducted. No significant differences were found between any video pairs (IG: -0.01, IB: 0.03, $t(21) = -0.36$, $p = .72$; IG: -0.01, OG: 0.02, $t(21) = -0.38$, $p = .71$; IB: 0.03, OB: -0.04, *t(21)* = 0.64, *p* = .53; OG: 0.02, OB: -0.04, *t(21)* = 0.42, *p* = .68). Given that ingroup good and outgroup bad videos were reacted to more quickly than ingroup bad and outgroup good videos, we hypothesized that the average response time to ingroup good/outgroup bad would be faster than the average response time to ingroup good and outgroup bad videos. Contrary to our hypotheses, participants were not significantly faster responding to ingroup good/outgroup bad videos than they were to ingroup bad/outgroup good videos (*t(21)* = -0.76, *p* = .46).

Results: Adult

Trends in reaction times

The trend in RTs for the Adult participants followed a pattern similar to that of the data from Experiment 1 child participants. Ingroup good and outgroup bad videos were responded to equally quickly, and ingroup bad and outgroup good videos were responded to equally slowly (see Figure 2).

Figure 2. Mean z-scores for Experiment 1 adult reaction times. SE values for Adults were 0.23 (IG), 0.17 (IB, OG), and 0.19 (OB).



Effects of group and action

A Repeated Measures ANOVA revealed a significant interaction between group and action (*F(1, 18)* = 4.35, *p* = .05). There were no significant effects of group (*F(1, 18)* < 0.00, *p* =.98) or action (*F(1, 18)* < 0.00, *p* = 1.00).

Differences between video pairs

The paired-samples t-test revealed no significant differences between the ingroup videos, nor between the bad videos (IG: -0.10, IB: 0.10, *t(18)* = -1.08, *p* = .29; IB: 0.10, OB: -0.10, *t(18)* = 1.61, *p* = .13). The differences between videos in the good pair and the outgroup pair were marginally different (IG: -0.10, OG: 0.10, *t(18)* = -1.86, *p* = .08; OG: 0.10, OB: -0.10, *t(18)* = 1.91, *p* = .08). Consistent with our hypotheses, participants were significantly faster responding to ingroup good/outgroup bad videos than they were to ingroup bad/outgroup good videos (t = -2.44, p = .02).

Discussion

In Experiment 1, participants were placed in minimal social groups and asked to make rapid decisions about whether an action was good or bad. The data trends for both children and adults followed our predicted directions that participants would respond more quickly to consistent videos (ingroup good and outgroup bad) than inconsistent videos (ingroup bad and outgroup good). Because they were not statistically significant, it is unknown whether these trends would hold up for children given more data. It is also impossible to tell whether they represent ingroup favoritism or outgroup discrimination.

For children, no significant overall effect of group or action was found, nor any interaction between those two factors. That is, across all videos, the time children took to decide whether an action was good or bad did not depend upon whether they were seeing good or bad actions or whether they were seeing their own or another group performing these actions. Furthermore, no significant interaction between group and action was found, meaning that group membership had no detectable influence on reactions to good and bad videos within that group, and vice versa.

Unlike children, adults reacted to consistent conditions equally quickly and to the same degree away from the average as inconsistent conditions, which were also reacted to equally quickly (or slowly). Adults did show a significant interaction between group and action. Though group membership or type of action alone did not alter adults' quickness to react to the videos, the interplay between these two conditions did significantly affect their reaction times. That is, adults exhibited an overall difference in response times to good and bad actions as a function of group membership.

Perhaps adults and children act differently in minimal groups, so that adults form biases immediately after being placed in a minimal group, but children need more induction and interaction with the groups for them to form measurable biases. This would help explain the disparate findings in minimal group literature, where some studies find that simply being placed in a minimal group is enough to produce intergroup biases, and others find that more group induction – for example, a competitive situation between the two groups, or functional use of group labels by an authority figure – is necessary to produce these biases (Ashburn-Nardo, et al., 2001; Bigler, et al., 1997; Bigler, et al., 2001; Otten & Wentura, 1999; Rhodes, 2009).

Experiment 2

The lack of significant results for children in Experiment 1 prompted the creation of Experiment 2, which sought to both increase group affiliation and measure sensitivity to group membership. Accordingly, Experiment 2 added two components to the study, which was otherwise identical to Experiment 1: a competitive group prime and an explicit measure of group preference. The purpose of the competitive prime was to encourage group differentiation in the participants' minds, in order to see whether more group induction would result in statistically significant reaction time differences between the four conditions (ingroup good, ingroup bad, outgroup good, outgroup bad). The purpose of the explicit measure was to test for the possibility that the lack of significant results in Experiment 1 was due to children's indifference to their assigned group membership.

Method

Participants

There were 20 participants (13 males; mean age 6;10) in Experiment 2. Participants were recruited through a database of local families and through local kindergartens and elementary schools. Testing occurred in the Wesleyan Cognitive Development Laboratory and at local schools.

Stimuli

Stimuli for Experiment 2 included all of the stimuli for Experiment 1, with the addition of the following items.

Prime pictures. The program Paint was used to create simple drawings for five scenarios: a sand castle building contest, tug-of-war, a carnival booth, kickball, and a cafeteria table. The pictures did not contain red or blue colors, and any people in the images were black stick figures. The sand castle picture depicted a yellow sand castle with a green flag; the tug-of-war picture presented two people pulling opposite ends of a brown rope; the carnival booth showed a simple booth outline with many differently colored (but not blue or red) balloons, and a person throwing a dart at the balloons; the kickball picture showed one person kicking a green kickball; and the cafeteria table picture depicted a simple black round table with a chair on either side.

Crystals. Crystals were made of bluish-purple translucent plastic. They were contained in a purple bowl and accompanied by two squares of paper, one red and one blue.

Procedure

The procedure for Experiment 2 included the procedure for Experiment 1, with the following additions.

After participants were placed into the red or blue group, but before they viewed the test videos, participants were shown the competition-inducing pictures as they listened to the accompanying scenarios. Short vignettes eliciting ideas of competition between the red and blue groups accompanied the pictures (see Appendix 4). Although the idea of competition was promoted (by saying, for example, that only one group won the game), no indication was given about who won.

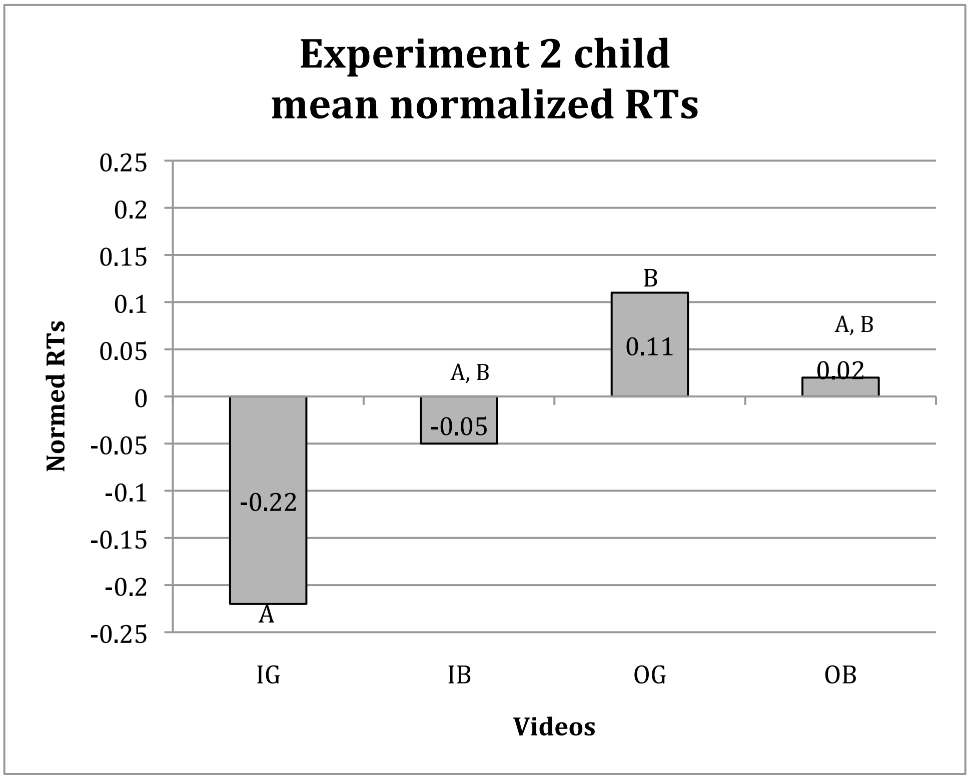
After viewing the test videos, the experimenter placed two squares of paper (red and blue) in front of the participant, then presented a small bowl filled with five crystals. The experimenter explained to the participant that he/she could give as many or as few crystals to the two groups as the participant wished (see Appendix 2 for script).

Results

Trends in reaction times

In Experiment 2, all ingroup videos were responded to more quickly than all outgroup videos (see Figure 3). Ingroup good videos were processed more quickly than average and outgroup bad videos were processed more slowly than average, but the ingroup and outgroup bad RTs fell in different directions from the mean than in Experiment 1.

Figure 3. Mean z-scores for Experiment 2 child reaction times Note: SE values for Experiment 2 were 0.20 (IG, OB), 0.19 (OG), and 0.21 (IB).



Effects of group and action

A Repeated Measures ANOVA revealed a marginal effect of group (*F(1, 19)* = 4.11, *p* = .06), and no significant effect of action (*F(1, 19)* = 0.10, *p* = .57) or interaction between group and action (*F(1, 19)* = 2.61, *p* = .12).

Differences between video pairs

Participants responded significantly faster to ingroup good than to outgroup good videos (IG: -0.22, OG: 0.11, *t(19)* = -3.12, *p* < .01). No other pairs showed significant differences (IG: -0.22, IB: -0.05, *t(19)* = -1.41, *p* = .18; IB:

-0.05, OB: 0.02, *t(19)* = 0.62, *p* = .54; OG: 0.11, OB: 0.02, *t(19)* = -0.42, *p* = .67). Participants were marginally faster responding to the average of the ingroup good/outgroup bad videos than they were to the average of the ingroup bad/outgroup good videos (*t* = -1.93, *p* = .07).

Crystal distribution

At the end Experiment 2, child participants were allotted five crystals to distribute between the red and blue groups. They distributed the crystals in the following way (n=20): 60% participants gave more crystals to their own group than to the other group, 30% participants gave the same number of crystals to both groups, and 10% participants gave more crystals to the other group than to their own group.

Of the females in Experiment 2, 38% (n=8) showed no ingroup bias (distributing either the same to both groups or more to the outgroup). Of the males in Experiment 2, 42% (n=12) showed no ingroup bias. Thus, sex was not a crucial factor in explicit measures of group preference, $\chi^2(1, 20) = 0.08$, $p = .78$.

Age, however, was a factor in resource allocation. A negative correlation existed between age and the proportion of crystals given to the ingroup in relation to all crystals available (-.47, *p* = .04), as well as between age an the proportion given to the ingroup relative to the total given (-.46, *p* = .06). However, for the ratio of outgroup: all and outgroup: total given, the correlations between the proportions and age are not significant. This suggests that as children age, they allocate less to their ingroup, but they do not allocate more to the outgroup, maintaining either absolute ingroup advantage or at least equality.

Of the children who gave more crystals to their own group, 83% (n=12) distributed in a 3:2 ratio. One child gave 4:1 and one gave 5:0. Thus, even when children were giving more crystals to their own group, they were only giving the minimum amount more that was necessary to produce a discrepancy.

Results: Combined data

Because the videos and reaction time measures were identical for Experiments 1 and 2, the two experiments were analyzed together to provide more power for the analyses and to test for significant effects across all data.

Trends in reaction times

The trends in mean RTs for combined data were as follows: ingroup good videos were processed the most quickly (-0.09), followed by outgroup bad (-0.02), ingroup bad (0.00) and outgroup good (0.03). Standard errors ranged between 0.13 and 0.14.

Effects of group and action

A Repeated Measures ANOVA revealed no significant effect of group ($F(1, 41) = 1.85$, $p = .18$), action ($F(1, 41) = 0.07$, $p = .79$) or interaction between group and action ($F(1, 41) = 1.62$, $p = .21$).

Differences between video pairs

As in Experiment 2, participants responded significantly faster to ingroup good than to outgroup good videos (IG: -0.09, OG: 0.03, $t(41) = -2.41$, $p = .02$). No other pairs showed significant differences (IG: -0.09, IB: 0.00, $t(41) = -0.90$, $p = .18$; IB: 0.00, OB: -0.02, $t(41) = -0.15$, $p = .88$; OG: 0.03, OB: -0.02, $t(41) = 0.42$, $p = .68$). Analyses separated by sex revealed that male participants responded with a marginal difference to the ingroup pair (*t(23)* = -1.99, *p* = .06). Contrary to our hypotheses, participants were not faster responding to ingroup good/outgroup bad videos than they were to ingroup bad/outgroup good videos (*t* = -1.27, *p* = .21).

Discussion

While the mean normalized RTs in Experiment 1 trended in the predicted direction, there were no significant effects of or interactions between group and action for either the child or adult participants, nor were there any significant differences between pairs of videos. Experiment 2 added more group induction in an attempt to elicit significant implicit biases.

The trends in Experiment 2 followed predicted patterns, and were stronger than Experiment 1. Again, there were no significant effects of or interactions between group and action. However, there was a significant difference in the good video pair. This finding held up when data from Experiment 1 and 2 were combined. This means that children responded significantly faster to good actions when their own group was performing the action than when the other group was performing it, but they did not respond differently to bad actions based on group. Because this finding demonstrates that detectable implicit biases exist, the lack of other statistically significant findings in the implicit measure suggests a possible lack of power or a small effect size.

Experiment 2 also added an explicit measure to determine whether any bias existed that was not being captured by the implicit measure. While there were few or no significant findings for the implicit measures, the explicit measure did find that children were biased toward their ingroup, confirming that children were sensitive to their group membership. This explicit this bias decreased significantly as a factor of age.

General Discussion

The automaticity of behavior evaluation based on group membership

In Experiment 1, minimal group membership was sufficient to significantly affect how quickly adults processed good and bad actions. In contrast, minimal group membership did not significantly affect the time children took to decide whether an action was good or bad. However, children's reaction times trended in predicted directions, implying that they were factoring group membership into their judgments at some level.

In an attempt to strengthen these trends, Experiment 2 increased group affiliation by adding a competitive prime. This had the effect of producing larger-scale trends, which included a significant difference in reactions to the good videos, with ingroup good videos reacted to more quickly than outgroup good videos. Combined data from the two experiments also found a significantly different response to good videos based on group. These findings indicate that participants held stronger expectations about which group members would perform good actions than about which would perform bad actions. Specifically, participants expected good actions from ingroup members and were surprised to see outgroup members being good, but did not hold expectations about which group would perform bad actions.

Any trend toward implicit favoritism or discrimination present in childhood leveled out by adulthood, where the data revealed equivalent levels of ingroup favoritism and outgroup discrimination. In Experiments 1 and 2, children responded (insignificantly) more strongly to one of the consistent conditions, and to one of the inconsistent conditions. With more power, these trends would reveal discrimination or favoritism. In Experiment 1, for example, children reacted to the outgroup bad videos the most quickly of the consistent conditions, and to the outgroup good videos the most quickly of the inconsistent conditions. If these trends were significant, that would point to more outgroup discrimination than ingroup favoritism. For adults, though, seeing an ingroup member perform a good action was no more contradictory to adults' expectations than was seeing an outgroup member perform a bad action, nor was seeing an outgroup member performing a good action more contradictory to expectations than seeing an ingroup member performing a bad action. Adults, therefore, exhibited equivalent amounts of ingroup favoritism and outgroup discrimination.

For adults, the same action performed by the same actor was reacted to differently based on whether the actor was wearing an identical or different t-shirt as the participant. The objectiveness of the action was skewed by the participants' own biases, which affected how quickly that person processed the perceived action. In relation to real social groups, one reason it would be helpful to interpret actions as either stereotype reinforcements or as exceptions to the rule would be for maintaining cognitive dissonance and justifying (implicitly or explicitly) a preference for one's own group over another group in the absence of objective justifications.

Sex differences

It is of interest to note that males responded with a marginal difference to ingroup good and ingroup bad videos, whereas females did not. Given that males in the experiment were significantly older than females ($p < .05$), it is quite likely that any discrepancy was actually due not to a sex difference but to an age difference. To confirm this, future studies would need to test an older female cohort and a younger male cohort. More data could shed light on the possibility of sex differences in reactions to ingroups and outgroups.

If the marginal difference between males' responses to ingroup good and ingroup bad videos were to gain significance with more data, while an equivalent-age female population's responses were to remain insignificant, it would suggest that males are predisposed to favor their own group more than are females. Supporting a possible sex difference is the finding boys as young as 6 have been shown to hold implicit views that differ from their explicit views (Baron & Banaji, 2006). This does not mean that girls cannot hold discrepant views as well; boys may just be more biased than girls at this age.

Also, compared to girls, boys are more opposed to disadvantageous inequality for an ingroup member than for an outgroup member (Fehr, et al., 2008). That is, they allocate in a way that maintains their own relative advantage over an outgroup member, but is maximally advantageous to both parties if the other person is an ingroup member. Again, this does not mean that girls allocate in a more disadvantageous manner, only that they may be less biased or more egalitarian to begin with.

If males', but not females', responses to ingroup good and ingroup bad videos were to gain significance with more data (and controlled ages), it would suggest that males are predisposed to favor their own group more than are females. Significantly faster reaction times to ingroup good than ingroup bad videos would mean that males expect to see an ingroup member perform a good action, and conversely find it hard to conceptualize their own group members acting in a negative manner. This would suggest that males are disposed to immediately generate positive views of their own groups (even if these are minimal and irrelevant). An early sex difference in implicit bias strength is supported by current research, but further studies would be necessary to investigate the reason behind this difference.

Explicit biases

Because the prime and the explicit measure were added at the same time, it is impossible to tell whether the prime had any confounding effect on results of the explicit measure. Regardless, the explicit measure was a necessary addition to at least make sure that even with the added group induction, participants were not still indifferent to explicit group preferences. While initially incorporated as a simple sensitivity measure, the explicit measure produced interesting findings on relationships between explicit and implicit biases in development and expression.

The fact that more than half of children gave to their own group demonstrated that the task was eliciting sensitivity to group membership. A preliminary analysis of the data found that the average age for distributing more to one's own group was 6;4, significantly lower than the average age for allocating equally, 8;5, prompting a further investigation of the relationship between age and allocations.

Age was a significant factor in determining how children shared with the two groups. A negative correlation was found between age and the proportion of crystals given to the ingroup in relation to the total crystals available. Yet while children were sharing less of their total resources with their own group as they aged, they were not sharing more with their outgroup. In this experiment, then, the decreased explicit bias in older children seems to reflect an attempt to monitor ingroup favoritism, rather than an attempt to decrease outgroup discrimination.

It is difficult to determine whether a decrease in ingroup favoritism was the same as a shift toward egalitarianism, or whether this was just an effect of the limited number of crystals available. For example, say that a child with 5 crystals decides to distribute three crystals to the ingroup and two to the outgroup. The only way that the child can lessen this bias in the future (without giving more overall to the outgroup) is to distribute equally. If instead participants were allotted 50 crystals to distribute, younger children could still distribute the same proportions, giving 30 to the ingroup and 20 to the outgroup. Older children, however, would now have a chance to distribute all of the resources in a way that was more egalitarian but still not completely equal, such as 28:22. Future experiments could include more crystals to examine patterns of sharing that arise with the possibility of distributing in finer ratios.

The age-related decrease in unequal resource allocation is supported by research showing that most 7-8 year-olds prefer resource allocations that remove advantageous or disadvantageous inequality (Fehr, et al., 2008). One possible reason that children start to monitor their explicit preferences as they age is that they are beginning to care that their preferences are socially acceptable (Baron & Banaji, 2006; Fehr, et al., 2008). As Fehr et al. (2008) state, "If older children care about what anonymous others think about them, they may be more prone to behave in a normatively appropriate way because – owing to their age – they are more likely to have theory-of-mind and perspective-taking abilities" (p. 1082).

Even though younger children were generally giving more crystals to their own group, they were only giving the minimum amount more that was necessary to produce a discrepancy. This suggests that young children are either only slightly biased toward the ingroup, or are already beginning to monitor their explicit biases. Again, because the explicit task was added as a measure of group sensitivity rather than as a controlled assessment of explicit biases, task manipulation and a larger age range are needed to further investigate the development and expression of implicit and explicit biases.

The relation between implicit and explicit biases

The fact that explicit biases were detectable but implicit biases were only trending in the predicted directions suggests that in young children, explicit biases develop might before implicit biases. It is possible, then, that explicit biases actually contribute to the development of implicit biases, rather than implicit biases developing first. The order of development of implicit and explicit biases has not been directly addressed by the intergroup-bias-development literature, so more studies are needed to determine the exact pattern of bias development (Dunham, Baron, & Banaji, 2008).

Baron and Banaji (2006) raised the question of whether implicit-explicit dissociations are even possible in children younger than age 6, whose conscious and less conscious attitudes may be more unified than they are in older children and adults, due to the fact that young children may not have the desire or capacity to monitor their explicit biases. The current findings seem to show that it is indeed possible to hold discrepant implicit and explicit attitudes even earlier than age 6; while 5- and 6-year-olds in this study showed ingroup preference on the explicit measure, there were no significant biases revealed by the implicit measure.

Given that children's explicit biases may not reflect any deep-seated nonconscious preferences, findings from other minimal group studies supporting relatively quick bias development within the paradigm may be not be as applicable to real life situations as they seem. That is, though children may in fact develop explicit biases quickly, these biases do not necessarily enter as quickly into children's reasoning or decision-making, and are therefore somewhat inconsequential.

From a social psychological standpoint, it is reassuring that most of the implicit measure results were not significant. Had they been, it would have meant that all it took to produce group biases was to be placed in a group. No prior experiences or third-party opinions would have factored into those biases. Given the amount of group induction, stereotypes, and experiences that factor into real life social groups, this would have been a troublesome finding for people working on conflict resolution, group alliance promotion, or stereotype reduction.

It is also promising that over a third of participants did not express an explicit ingroup preference on the explicit measure. This means that even if children felt a liking toward their own group, they were conscious of the fact that there was really no reason to like their group more, and they express this understanding through more equal resource allocation. This bias-reducing mentality has been shown to significantly decrease implicit biases in adults. In one experiment, those who implicitly held negative attitudes towards prejudice showed less influence of implicit stereotypes on an automatic-discrimination task (Glaser & Knowles, 2008). Implicit intergroup biases, therefore, can be countered by biases against prejudice. Future studies are needed to more fully examine how bias is expressed through resource allocation compared to automatic judgments.

Bayesian reasoning

We hypothesized that the experiment would reveal Bayesian reasoning. While the lack of many significant results makes it difficult to tell, it seems that participants were trending toward Bayesian reasoning by taking into account group membership when making decisions. While not necessarily predicted by Bayesian reasoning, the trends toward adult-like levels of implicit biases represent one possible manifestation of this reasoning.

Bayesian reasoning does not say how biases will enter into reasoning, just that they may be incorporated in some way. Traditional statistical reasoning would predict that all conditions be responded to the same, since there is no objective difference between red-shirted or blue-shirted people performing the same action, and therefore an equally likely probability that a good or bad action will be performed by either group. The fact that shirt color did have an effect on how participants perceived behaviors indicates that some type of reasoning beyond traditional probability reasoning was occurring.

Limitations and future directions

Because a similar minimal group reaction time experiment for children (or adults) had never been conducted before, it was uncertain how much group induction would be necessary to produce significant results. Future experiments could manipulate group membership to determine exactly how much group induction and/or competition is necessary to produce significant implicit group biases. The blue-eyed/brown-eyed experiment, where one group at a time was explicitly favored while the other was explicitly discriminated against, provided an example of sufficient group manipulation for producing intergroup biases (Stewart, et al., 1998). Future experiments could determine the point between the current experiment and the blue/brown experiment where significant implicit biases are elicited.

Future versions of this experiment could also include more precise stimuli videos. These videos were made so that each lasted a specific number of seconds depending on which activity was being shown. In the future, a brief time limit could be set following the end of the action (whenever that is determined to be) to better control for children who wait until the end of the videos to respond. Also, the videos could be cut immediately after the action and reaction times could be recorded only after videos so that it is clearer to children when to respond.

Editing the videos to end at a specific time after completion of the action would still leave the problem of objectively deciding when the action ends. For example, in the bad-puzzle video, one actor flips over another actor's completed puzzle, to which the second actor exclaims, "Hey, I spent all day putting that puzzle together!" The exclamation at the end was added to make the second actor's feelings objectively clear. This potentially leaves two points at which to decide whether the action is good or bad – at the end of the action or the end of the verbal response to the action.

Many children found it obvious after the first actor flips the puzzle that this was a bad action, and research confirms that very young children can sympathize with a victim even in the absence of overt emotional signals (Vaish, Carpenter, & Tomasello, 2009). This scenario could be improved upon by showing only nonverbal videos to create a situation where there is only one point where a decision could be made. The puzzle video, for example, could end a certain amount of time after the puzzle flip and responses could be recorded only after the action was complete.

The analyses attempted to control for those who waited until the ends of the videos by performing analyses without 5-year-olds and with only 5-year-olds, who seemed anecdotally to have a lesser understanding of the task than older children or to be more likely to wait until the end of each video to respond, even when it was apparent that they knew their response beforehand, and were reminded that they could answer before the videos ended. Separate analyses, however, did not find any significant differences in the data without 5-year-olds, and 5-year-olds did not differ significantly from older children. It seems then that the younger children were responding in the same patterns as older children, and may have simply taken longer to process and respond to the videos. In the future, though, it would be helpful to control for the tendency to wait for videos to end. As stated previously, this could be accomplished by ending the videos more quickly after the action ends and by only capturing reaction times after the end of each video.

Finally, an important improvement would be to carry out this experiment with full counterbalancing. Due to time constraints and the nature of participant recruitment, and in a compromise between attaining the highest possible power and making sure that every order was tested, this experiment did not carry out counterbalancing to the fullest extent possible. There were four different video sets included in testing, but ideally there would have been more orders included in testing. Because of this imperfect counterbalancing, it happened that certain videos appeared more at the ends and other videos more at the middle of the video sets.

Combined with imperfect counterbalancing was the problem that repeating situations could have led children to recognize the situation and even type of action earlier in the next similar video, so reaction times often sped up over the course of the experiment. Therefore, since some videos were played disproportionately at the beginning or ending of testing, the increasingly fast reaction times could lead to an effect of item (specific videos having faster or slower reaction times). Exhaustive counter-balancing is therefore a suggestion for future studies in case there is some reason to believe that there is an effect of video, in which case this could be reliably tested. Also, increasingly faster reaction times would have no bearing on the averaged results if videos were counterbalanced completely and if participants were perfectly distributed across conditions. In our experiment, and with only 20-23 participants per condition, this was not the case. Normed scores reconciled this problem to some extent, but future studies may focus on analyses of raw scores, in which case perfect counterbalancing, video editing and participant recruitment is necessary.

Future versions of this experiment should maintain or improve upon the current experiment's integrity in attempting to eliminate confounding factors from the videos. The videos in the current experiment were clearly good or bad, and a long practice period ensured that children understood how to respond correctly to the videos. Actor ethnicity, gender, and role in the videos were controlled for, along with performance space, lighting, and shirt color intensity.

One question left unanswered by the current experiment is: How quickly does explicit bias development? It would be of interest in future experiments to determine whether explicit biases develop immediately after placement into a minimal group, or whether watching both groups perform good and bad actions contributes to the development of this bias. To test this, some participants' explicit group preferences could be assessed immediately after group designation. However, this manipulation could only be performed by some of the participants, as any subsequent implicit measure would have to be disregarded due to the propensity for explicit measures to bring conscious awareness to the target manipulation and confound the results of implicit measures.

Another major direction for further research on the topic is to determine whether the group bias detected in minimal group situations is due to ingroup favoritism, outgroup discrimination, or both (and if both, the relative levels of each). Answering this question has practical implications in reducing bias in children. In order to reduce bias with real social groups, knowing which direction the bias comes from would be very helpful for focusing efforts. For example, if in a given situation it is determined that there is much more ingroup favoritism than outgroup discrimination, then efforts could focus on highlighting how the outgroup shares some of the same valued characteristics of the ingroup. If instead ingroup favoritism is minimal, but outgroup discrimination is very high, positive examples of outgroup members or members' behavior could be used to place the outgroup in a more positive light. Multiple methods could be employed if both favoritism and discrimination are at work.

Finally, future experiments should attempt to determine how much group induction and/or association is necessary to produce different levels of implicit intergroup biases. Experiments could determine exactly how much competition invocation is necessary to produce significant effects of group or action. From there, it is essential to test how bias development differs depending on the nature of group differentiation (competition, functional use of labels by an authority figure, etc.).

Conclusion

The results of this experiment indicate that placement into minimal groups is not sufficient to produce implicit biases in children strong enough to affect simple decision-making, but that placement alone may be sufficient for eliciting these decision-affecting biases in adults. Giving children a competitive group prime increases their levels of bias. In adults this bias manifests itself equally in ingroup favoritism and outgroup discrimination, but with children it is difficult to tell. As they grow older, children likely become more aware of not only how their views compare with other peoples' views, but of how other people may judge the their views (Dunham, et al., 2008). Thus, while individuals cannot control their implicit group biases, with age they begin to control the outward expression of those biases.

Works Cited

Aboud, F. E. (1988). *Children and prejudice*. Cambridge, MA: Basil Blackwell.

Ashburn-Nardo, L., Voils, C. I., & Monteith, M. J. (2001). Implicit associations as the seeds of intergroup bias: How easily do they take root? *Journal of Personality and Social Psychology, 81*(5), 789-799.

Baron, A. S. & Banaji, M. R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science, 17*(1), 53-58.

Baron, A. S. Dunham, Y., Banaji, M. R., Carey, S. (under review). Constraints on social categorization.

Bigler, R. S., Jones, L. C., & Lobliner, D. B. (1997). Social categorization and the formation of intergroup attitudes in children. *Child Development, 68*(3), 530-543.

Bigler, R. S., Spears Brown, C., & Markell, M. (2001). When groups are not created equal: Effects of group status on the formation of intergroup attitudes in children. *Child Development, 72*(4), 1151-1162.

Diesendruck, G., & haLevi, H. (2006). The role of language, appearance, and culture in children's social category-based induction. *Child Development, 77,* 539-553.

Dunham, Y., Baron, A. S., & Banaji, M. R. (2008). The development of implicit intergroup cognition. *Trends in Cognitive Science, 12*(7), 248-253.

Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature, 454*(7208), 1079-1083.

Gaertner, L., & Insko, C. A. (2000). Intergroup discrimination in the minimal group paradigm: Categorization, reciprocation, or fear? *Journal of Personality and Social Psychology, 79*(1), 77-94.

Gallistel, C. R., & King, A. P. (2009). Bayesian updating. *Memory and the computational brain: Why cognitive science will transform neuroscience* (pp. 27-42). West Sussex, UK: Wiley-Blackwell.

Glaser, J., & Knowles, E. D. (2008). Implicit motivation to control prejudice. *Social Cognition, 26*(4), 401-419.

Goldinger, S. D., He, Y., & Papesh, M. H. (2009). Deficits in cross-race face learning: Insights from eye movements and pupillometry. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 35(5), 1105-1122.

Grieve, P. G., & Hogg, M. A. (1999). Subjective uncertainty and intergroup discrimination in the minimal group situation. *Personality and Social Psychology Bulletin, 25*(8), 926-940.

Hertel, G., & Kerr, N. L. (2001). Priming in-group favoritism: The impact of normative scripts in the minimal group paradigm. *Journal of Experimental Social Psychology, 37*(4), 316-324.

Kinzler, K. D., Shutts, K., & Correll, J. (in press). Priorities in social categories. *European Journal of Social Psychology: Special Issue: Children's Intergroup Attitudes*.

Lemyre, L., & Smith, P. M. (1985). Intergroup discrimination and self-esteem in the minimal group paradigm. *Journal of Personality and Social Psychology, 49*(3), 660-670.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The implicit association test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Automatic processes in social thinking and behavior* (pp. 266-286). Philadelphia: Psychology Press.

Otten, S., & Wentura, D. (1999). About the impact of automaticity in the Minimal Group Paradigm: Evidence from affective priming tasks. *European Journal of Social Psychology, 29*(8), 1049-1071.

Ray, D. G., Way, N., & Hamilton, D. L. (accepted manuscript, 2009). Crossed-categorization, evaluation, and face recognition. *Journal of Experimental Social Psychology*.

Rhodes, M. (2009, October). Preschoolers' representations of coalition alliances. In J. Wooley (Chair), *Phylogenetic and ontogenetic consequences of group membership for intergroup cognition*. Symposium conducted at the Cognitive Development Society VI Biennial Meeting, San Antonio, TX.

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science*, *28*(3), 303-333.

Sherif, M., Harvey, O. J., White, B. J., Hood, W. R., & Sherif, C. W. 1954/1961. *Study of positive and negative intergroup attitudes between experimentally produced groups: Robbers Cave study.* Norman, Oklahoma: University Book Exchange, 1961.

Stewart, T. L., Doan, K. A., Gingrich, B. E., & Smith, E. R. (1998). Do the "eyes" have it? A program evaluation of Jane Elliott's "blue eyes/brown eyes" diversity training exercise. *Journal of Applied Social Psychology, 33*, 1898-1921.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behavior. *European Journal of Social Psychology, 1*(2), 149-178.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. Austin (Eds.), *Psychology of intergroup relations* (pp. 7-24). Chicago: Nelson-Hall.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. New York: Blackwell.

Vaish, A., Carpenter, M., & Tomasello, M. (2009). Sympathy through affective perspective taking and its relation to prosocial behavior in toddlers. *Developmental Psychology*, *45*(2), 534-543.

Xu, F. (in press). Rational statistical inference and cognitive development. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Foundations and the future, vol. 3* (pp. 1-39). Oxford University Press.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences, 105*(13), 5012-5015.

Appendix 1.

Experiment 1 script

We're going to play a game today. Do you want to play a game? Great! The first thing we're going to do is practice using the computer. That way we'll know how to use it when we play our game. Are you ready to practice with the computer?

Ok. You're going to see some videos. In each video, someone will be doing something good or something bad. If you see someone doing something good, I want you to press this [smiley face] key, and if you see someone doing something bad, I want you to press this [frown face] key. I want you to press it as soon as you know whether it's good or bad. And only press it once per video.

So tell me, if you see someone doing something good, which key will you press? And if you see someone doing something bad, which key will you press? And how many times can you press the key?

Show trial videos

Great job! I'm going to put the computer over here for a little bit. We'll get to use it again in a few minutes. Ok, now we are ready to play our game! In this game, you are in the [red/blue] group. Can you put on this t-shirt for me? Thank you! And can you put on these wristbands for me? Thank you! So which group are you in? Now I'm going to show you some pictures. I want you to tell me if the person in each picture is in your group or the other group.

Show pictures

Ok, now we are ready to use the computer again. On the computer, you're going to see some videos just like before. Again, I want you to press this key if you see someone doing something good, this key if you see someone doing something bad. Remember, press it as soon as you know whether it's good or bad, but only press it once per video.

Show test videos

Appendix 2.

Experiment 2 script
Same as Experiment 1, with the following additions:
Show pictures

Now I'm going to read you some stories about the blue and the red groups. [See Appendix 4 for stories]

…

Show test videos

I have five crystals here. These crystals are really nice and everyone in the blue and red groups likes them. You get to give these crystals away to the groups. You can give only some away and keep some on the bowl, or you can give them all away. You can give more to one group, or the same to both groups. It's up to you! Just put however many crystals you want to give to each group on the pieces of paper.

Appendix 3.

Training videos description:

Two female actors acted for all of the candy videos and a male and a female actor acted for all of the drawing videos. In the good candy video, the two actors sat at a table, and one had a pile of candy in front of her. She asked the person with no candy pile, “Would you like some candy?” and the other person responded, “Yes, thank you.” The “good” person then handed some candy to the other person. In the bad candy video, one person grabbed some candy from the other person’s pile. With a sad expression, the other person exclaimed, “Hey! That was mine.”

In the good drawing video, both actors were seen drawing across from each other at a table. One person then looked over at the other person’s paper and said, “That’s a pretty picture,” to which the second person responded, “Thank you.” In the bad drawing video, both actors were drawing across from each other at a table. Then, one person reached over and scribbled on the second person’s drawing. The second person exclaimed, “Hey, you messed up my drawing!”

Test videos description:

The neutral person sat on the right side of the screen. In the good puzzle video, the target person entered from the left of the screen, sat down across from the neutral person, and asked, “Would you like some help putting that puzzle together?” The neutral person smiled and replied, “Sure. Thanks!” In the bad puzzle video, the neutral person was shown putting the last piece in the puzzle, after which the target person entered, sat down, and overturned the puzzle, saying, “My turn!” With a pained expression, the neutral person replied, “Hey, I spent all day putting that puzzle together!”

In the fall videos, a neutral person walked in from the left side of the screen and fell down, exclaiming, “Ow!” In the good fall video, the target person entered from the right side of the screen and extended his/her hand to the fallen person, saying, “Here, let me help you.” The fallen person responded, “Thank you.” In the bad fall video, after the person fell the target entered from the right, pointed to the fallen person and laughed, “Ha-ha!” and continued to walk off the left of the screen.

Appendix 4.

Experiment 2 stories

Cafeteria
In school today, the red and blue groups both wanted to sit at this cafeteria table. But there were not enough chairs for both groups, so only one group got to sit at the table.

Sand castle
The blue and red groups were at the beach and there was a sand castle building contest. Only one of the groups built the best sand castle.

Carnival
The blue and red groups are at the carnival. They are throwing darts at balloons, and if they hit the balloons they win a prize. Only one group will win a prize.

Tug of war
At recess, the blue and red groups were playing tug of war. One group pulled the rope the hardest to win tug of war.

Kickball
At recess, the blue and red groups were playing kickball. One of the groups won win the kickball game.