The Influence of Group Membership on Preschoolers' Selective Trust

by

Kyle MacDonald
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

Acknowledgements

First, I would like to thank my advisor Hilary Barth for her invaluable guidance throughout this entire project. Thanks to my co-advisor Mariah Schug for all her help designing this experiment and for her advice during the writing process. Thanks to my family: my parents Richard MacDonald and Morag MacDonald for providing an endless supply of inspiration and motivation, and my sister Kaitlin MacDonald for her support and for being such a great friend. Thanks to Jennifer Garcia, Rebecca Rieger, and Elise Herrig for their help scheduling families and for making the lab an enjoyable work environment. Thanks to my fellow Senior Thesis writers: Dominic Gibson and Gwynne Hunter for providing company on this long journey. A special thanks to all the Wesleyan students who graciously volunteered to act in my videos, and to all the families who participated in my study. Finally, thanks to Ellika Healy for her moral support and for reading/editing this thesis multiple times while still managing to appear interested. I could not have done this without you.

Abstract

Is a speaker's group membership or a previous history of accuracy/inaccuracy a more powerful cue during children's evaluations of trustworthiness? Across two experiments, 4-year-old children's selective trust capabilities were altered by the presence of group membership. In Experiment 1, children were placed into one of two minimal groups (Blue Group v. Red Group) and watched films establishing one in-group member as reliable and another in-group member as unreliable across the following rates of accuracy and inaccuracy, respectively: 100% v. 0%, 100% v. 25%, 75% v. 25%, and 75% v. 0%. Next, children witnessed the two speakers provide conflicting novel labels for unfamiliar objects, and were asked to endorse one label over the other. Four-year-olds selectively trusted the more accurate informant in only two out of four conditions: 100% v. 0% and 75% v. 0%. In Experiment 2, children watched films in which out-group speakers were more reliable than in-group speakers across the four accuracy/inaccuracy ratios. Children were unable to selectively trust the more reliable out-group member in all four conditions. Thus, 4-year-olds' demonstrated ability to trust more reliable individuals is desensitized when evaluating in-group members and disoriented when comparing a reliable out-group speaker to an unreliable in-group speaker. These results suggest that group membership processing is privileged during young children's evaluations of others' trustworthiness.

Table of Contents

General Introduction

Humans enter a complex physical and social world equipped with a small number of separable systems of core knowledge (e.g., object, agents, number, and place representations), which provide a foundation for the learning of more complex skills (Spelke & Kinzler, 2007). However, despite the presence of these core knowledge systems, much of children's learning occurs via social transmission, and therefore inherently relies on the testimony of others. This dependent relationship is intrinsically problematic because the reliability of others is often not guaranteed. A given speaker might be either ignorantly incorrect or intentionally misleading when making assertive claims. Therefore, in order to become efficient learners, it is likely that humans develop a categorization mechanism that evaluates incoming information and decides whether to accept it as reliable, reject it as inaccurate, or suspend judgment until further evaluation (Faulkner, 2002). How does such a system develop and function? Do children test everything they learn against personal experience, or are they capable of tracking the reliability of certain individuals to make an educated decision about whom to trust?

Young children do in fact possess a mechanism that tracks, appraises, and draws on a speaker's past reliability when evaluating the credibility of their current testimony (Birch, Vauthier, & Bloom, 2007; Corriveau & Harris, 2009a; Corriveau & Harris, 2009b; Koenig & Harris, 2004; Koenig & Harris, 2005; Koenig, Clement, & Harris, 2004; Pasquini, Corriveau, Koenig, & Harris, 2007). Furthermore, there is a distinct progression of evaluative abilities, which begins with basic gaze-tracking skills displayed by 14- to 16-month-olds (Chow, Poulin-Dubois, & Lewis, 2008;

Poulin-Dubois, & Chow, 2009) and culminates in a nuanced version of selective trust that emerges in 4-year-olds. Specifically, 4-year-olds are able to (a) differentiate between a reliable and an unreliable speaker even when the difference between the two is slightly ambiguous (Pasquini et al., 2007), (b) reverse trust when a speaker is later determined to be unreliable (Scofield & Behrend, 2008), and (c) maintain trustworthy versus untrustworthy attributions over time (Corriveau & Harris, 2009a). Taken together, these findings demonstrate that young children's reliability tracking systems are quite sophisticated, and result in remarkably efficient evaluations of trustworthiness.

In the real world, however, young children are not limited to information concerning reliable versus unreliable output when deciding if a given speaker is trustworthy. In addition to this information, children use various social cues to help evaluate testimony. For instance, children monitor a speaker's confidence level, choosing to endorse a more confident speaker more often (Jaswal, 2004; Jaswal & Neely, 2006). Further, children observe and use bystander cues, more often endorsing a speaker's claim that is coupled with positive bystander reactions (e.g., nodding and smiling) (Fusaro & Harris, 2008).

If children take social cues such as a speaker's confidence and bystander reactions into consideration, what other variables might influence their reliability judgments and ensuing evaluations of trustworthiness? One interesting prospect is group membership. It has been well documented that humans have an aptitude and a propensity for social categorization along some dimension (e.g., race, gender, language) (Kelly, Quinn, Slater, Lee, Gibson, Smith, Liezhong & Pascalis, 2005;

Kinzler & Spelke, 2005; Kinzler, Shutts, & Correll, in press; Quinn, Yahr, Kuhn, Slater, & Pascalis, 2002). Moreover, this categorization is often accompanied by a preference for the similar in-group over the dissimilar out-group. For example, children (6- and 10-years-old) and adults have demonstrated implicit and explicit biases favoring their in-group based on the social category of race (Baron, & Banaji, 2006). Additionally, children demonstrate strong preferences for their own gender in-groups (Maccoby & Jacklin, 1987), and for native accents (Kinzler & Spelke, 2005; Kinzler, Shutts, DeJesus, & Spelke, 2009; Kinzler, Corriveau, & Harris, in press).

In-group favoritism, however, is not only limited to natural groups. A series of studies using the Minimal Group Paradigm (MGP), which places individuals into different categories based on arbitrary and seemingly meaningless characteristics (e.g., shirt color), demonstrate that the mere perception of belonging to two distinct groups is sufficient to engender in-group biases (see Brewer, 1979 for a review). For instance, in a classic study conducted by Tajfel and colleagues, members of two arbitrarily formed groups that had little contact or incentive to compete, still allocated more goods to in-group members and treated out-group members competitively (Tajfel, Flament, Billig, & Bundy, 1971). In sum, the research on both natural and minimal group bias demonstrate the ease with which children and adults categorize the social world, form *us* versus *them* attitudes, and develop an overall preference for their in-groups. Moreover, children will readily use these social categories to make psychological inferences about individual group members (Diesendruck & haLevi, 2006) and to dictate future interactions with both in- and out-group members (Kowalski & Lo, 2001).

In addition to supporting a general in-group preference, the act of social categorization results in specific and measurable consequences. These consequences include an increased perception of between-group differences and within-group similarity (Bigler, Jones, & Lobliner, 1997; Rothbart, Davis-Stitt, & Hill, 1997), an increased perception of out-group homogeneity (Park & Rothbart, 1982), and biased memory processes (Meissner & Brigham, 2001). Furthermore, group members may commit the "fundamental attribution error" when evaluating out-group behavior, disproportionately attributing negative out-group actions to their dispositions as opposed to the situation (Hamilton & Trolier, 1986). Thus, classifying individuals into disparate groups has the capacity to influence specific cognitive processes, making it an interesting potential confound for children's reliability tracking systems.

This thesis will attempt to address the potential influence of minimal group categorization and group bias on young children's evaluations of a speaker's trustworthiness. Will children choose to trust and subsequently learn from a reliable out-group member over an unreliable in-group member? Or will the cue of group membership prove too powerful to be overcome by a history of reliability? Efforts to provide insight into the relationship between group membership and trust have, for the most part, been limited to economic and management research conducted with adults. This work has shown mixed results. For example, in an examination of how social distance created with the Minimal Group Paradigm impacts trust, Buchan, Johnson, and Croson (2006) found that adults displayed strong in-group favoritism when deciding whom to trust. However, in a similar experiment, Guth, Vittoria Levati, and Ploner, (2007) found no effect of group membership on trust behavior.

Developmental research has demonstrated that children take a speaker's group membership into account when assessing their credibility. In the absence of prior reliability information, children will trust an adult over a child (Jaswal & Neely, 2006), a familiar adult over an unfamiliar adult (Corriveau & Harris, 2009b), and a native-accented speaker over a foreign-accented speaker (Kinzler, Corriveau, & Harris, in press). These studies, while demonstrating the potential influence of a speaker's social category on trust, have not directly addressed the interaction of prior reliability and in- or out- group status. Therefore, this thesis will investigate whether shared group membership or a previous history of accuracy takes precedence when children evaluate the trustworthiness of another individual.

Development of Children's Selective Trust

Infant Abilities

The development of selective trust begins early in life with infants' basic ability to track others' reliability and to differentiate a reliable from an unreliable individual. This capability serves as the foundation for the progression of children's skillful assessment of others' testimony. For example, Chow et al. (2008) recorded the gaze tracking behavior of 14-month-olds after they were exposed to reliability information about an actor's past looking behavior. The prior reliability of a looker was established with a training task in which infants saw an actor express excitement when looking into a container that either was empty (unreliable looker) or held a toy (reliable looker). Once looker reliability had been established, Chow and colleagues measured the gaze following behavior of infants in response to the same actor (i.e., a reliable or unreliable looker) looking either in front of or behind a barrier. Infants who were exposed to the reliable looker were more likely to follow his/her gaze to the out-of-sight target behind the barrier, although, when the target was in plain sight, infants in both the reliable and unreliable looker conditions performed equivalently. These results suggest that infants monitored prior reliability information and subsequently used that information by trusting a reliable looker's gaze, following it more often even when the target of that gaze was not visible to the infant.

Subsequent research provides further evidence in support of infants' rich reliability tracking abilities. Poulin-Dubois and Chow (2009) found that 16-month-old infants were able to successfully monitor the reliability of an individual's gaze and, interestingly, were able to use that information to evaluate behavior in a different

context. Looker reliability was first demonstrated using a training task (similar to Chow et al., 2008). After reliability had been established, infants' willingness to use this information was measured with a violation of expectation paradigm during a true-belief task. The true-belief task consisted of an experimenter hiding a cup in one of two boxes, and then searching for the cup in a location that was either consistent or inconsistent with her belief about where it was hidden. Infants in the reliable looker condition were surprised (looked longer) when a previously reliable looker demonstrated an incongruent true belief. In other words, children looked longer when a reliable individual searched in the wrong place for a toy to which he/she had prior perceptual access. Thus, infants were able to (a) appraise reliability, (b) encode the identity of a reliable person, and (c) make a cross-domain inference of knowledge based on past reliability information.

Infants' reliability monitoring, however, is not limited to looking behavior. Not only do infants attend less to unreliable lookers, but they also will respond questioningly when a speaker provides an inaccurate label for a familiar object. To demonstrate this, Koenig and Echols (2002) presented 16-month-old infants with familiar objects (e.g., chair, duck, cat, ball, and shoe), accompanied by either a correct or incorrect label. Infants in the False label condition looked longer at the *speaker* compared to the *object*, whereas infants in the True label condition displayed the opposite behavioral pattern – looking longer at the *object* compared to the *speaker*. In addition, when infants were confronted with a mismatch of label and object, they made explicit attempts to interrupt and correct the inaccurate speaker (see also Pea, 1982). Interestingly, in a series of follow-up studies, Koenig and Echols

found that this pattern was not replicated when the source of the false label was a non-human audio speaker, or when the human source of the correct label lacked perceptual access to the object. Taken together, these findings support a rich interpretation of infants' behavior as an attempt to understand others' knowledge states. If this were not the case, infants would have failed to demonstrate attention to characteristics of the speaker (e.g., perceptual access and human vs. non-human) (Koenig & Echols, 2002).

The research discussed thus far demonstrates that infants possess a range of speaker and reliability tracking capabilities such as (a) differentiation of true and false statements, (b) rejection of false claims, and (c) encoding of the identity of an unreliable speaker. However, this research does not address the scope of children's use of this information. For instance, do children create a trustworthiness profile of a given speaker, which involves a 'mentalistic attribution' (Harris, 2007)? Or do children simply evaluate a speaker's output (reliable vs. unreliable) without assessing his/her inner mental state (Nurmsoo & Robinson, 2009)? Additionally, will children consider prior evidence of reliability when deciding from whom to learn? Subsequent research with older children addresses these interesting questions.

Two-year-olds' Early Selective Trust

In an attempt to address the questions raised above, research on selective trust shifted from analyzing testimony that children could check against their own experience to measuring children's ability to assess the veracity of an unfamiliar

claim. This paradigm shift provides more direct evidence that young children display selective trust in specific individuals.

Even at the young age of two, children are able to display selective trust for reliable over unreliable speakers (Koenig & Woodward, 2007 as cited in Koenig & Harris, 2007). To demonstrate this ability, Koenig and Woodward presented one group of toddlers with an adult who provided correct labels for familiar objects, while a second group was presented with an adult who provided incorrect labels. Then, the adult provided novel labels for the test objects, and infants were asked to select the target object. Results showed that toddlers did not treat the novel labels provided by accurate and inaccurate speakers equally: they were less likely to endorse a previously inaccurate speaker's label for the target object. This finding suggests that by the age of two, children moderate their trust behavior based on an individual's demonstrated history of reliability.

Koenig and Woodward also collected a measure of vocabulary size in an attempt to address whether experience with language mediated trust abilities. There was an overall difference in performance between the high and the low vocabulary groups, but even children in the low group were able to selectively mistrust inaccurate speakers. Therefore, it appears that toddlers' decisions to trust or mistrust are not entirely reliant on prior language familiarity, but instead depend on experiences with specific individuals. This finding, combined with the infant research, supports the idea that young children are cognizant of specific behavior (i.e., prior reliability) and link it to specific individuals (i.e., evaluations of trustworthiness) (Koenig & Harris, 2007).

Preschoolers

Selective Trust. The majority of previous research on the selective trust capabilities of preschool aged children documents a distinct shift between the ages of 3 and 4. Specifically, a series of studies have shown that 3-year-olds, while capable of recognizing accurate versus inaccurate speakers, are incapable of using this information to predict future accurateness, direct questions towards, and endorse previously reliable speakers when learning novel words (Clement, Koenig, & Harris, 2004; Koenig, Clement, & Harris, 2004; Scofield & Behrend, 2008). However, certain studies have failed to produce significant age differences in similar reliability assessments, finding that 3-year-olds were able to use previous reliability information to assess trustworthiness in future interactions (Birch et al., 2007; Jaswal & Neely, 2006; Nurmsoo & Robinson, 2009). Regardless of the existence of a developmental shift in reliability tracking ability, by the age of 4, children are able to track a speaker's history of reliability and use this information when evaluating a speaker's future credibility.

Typically, selective trust has been measured with a standard reliability paradigm employed across numerous studies. For example, Koenig, Clement, and Harris (2004) presented 3- and 4-year-old children with video clips, which depicted a neutral actor asking two informants to label a familiar object that was placed on a table between the two informants. One speaker consistently labeled the familiar objects accurately (e.g., calling a ball "a ball"), while the other speaker consistently labeled the objects inaccurately (e.g., calling a ball "a shoe"). Children were then asked to explicitly identify the reliable or the unreliable speaker. Following these

familiarization trials, children saw the same actors provide novel labels (e.g., "toma" or "mido") for unfamiliar objects. As a measure of trust, children were asked what the unfamiliar object was called ("a toma or a mido?"). Results showed that both age groups performed well on the explicit judgment questions, successfully identifying the reliable or unreliable informant. Moreover, Koenig et al. (2004) found that the children who were most successful at the explicit judgment questions were more likely to endorse the novel label provided by the previously accurate informant.

In a series of follow-up studies, Koenig and Harris (2005) examined whether preschoolers would not only endorse previously accurate speakers but also turn to them to learn new information. These experiments involved a design and procedure similar to that of Koenig and Harris (2004), but they included a greater variety of test trials, which consisted of four types of questions: (a) explicit judgment, (b) prediction, (c) ask, and (d) endorsement. As in previous research, 3- and 4-year-olds performed well on the explicit judgment questions, demonstrating successful reliability tracking. However, only 4-year-olds were able to use that information to predict future assertions, seek information from the more accurate speaker, and endorse the more accurate speaker's claims (i.e., demonstrate selective trust). Furthermore, in an interesting manipulation, Experiment 1 pitted an *accurate* versus an *inaccurate* informant, whereas Experiment 2 pitted an *accurate* versus an *ignorant* informant. With this design change, 3-year-olds demonstrated selective trust abilities equivalent to 4-year-olds. Thus, 3-year-olds preferred an accurate informant to someone lacking in knowledge, but they did not prefer an accurate informant to an inaccurate informant. Taken together, these findings suggest that children

comprehend the utility of past reliability when determining the veracity of a speaker's future claims, but just how sophisticated is this mechanism?

Nuanced Selective Trust. In an attempt to address this question, Scofield and Behrend (2008) used a slightly modified version of the typical reliability paradigm to examine whether 3- and 4-year-old children could reverse trust and word mappings when a trusted speaker later proved unreliable. The key manipulation involved establishing speaker reliability after children had already formed an initial trusting preference. Children were presented with a 2-minute animated video in which two speakers entered the scene sequentially. The first speaker pointed to and labeled one of two unfamiliar objects (e.g., a shelving bracket) with a novel word (e.g., "This is a koba."). Children were then asked to choose which of the two unfamiliar objects best matches "a koba". After the initial endorsement (i.e., after children bestowed trust), the speaker re-entered and labeled familiar objects either correctly or incorrectly, thus establishing their reliability. Next, the second speaker entered and provided the same novel label ("koba") for the remaining unfamiliar object (e.g., a t-joint). This speaker then labeled the remaining familiar objects either reliably or unreliably (whichever the first speaker had not done). As a final measure, Scofield and Behrend asked children to choose which of the two novel objects was actually a "koba". Results showed that 4- but not 3-year-olds were capable of reversing trust and word mappings if an initially trusted speaker later proved unreliable.

The success of 4-year-olds on the task described above is a powerful demonstration of children's nuanced selective trust. In order to revise word mappings, children had to continuously monitor: (a) the trustworthiness of the original speaker,

(b) the original word, (c) the trustworthiness of the new speaker, and (d) the new word (Scofield & Behrend, 2008). In addition to simply reversing trust, children were able to maintain the trust revision over a time delay, displaying selective trust by choosing the reliable speaker's object when asked 24 hours later. In sum, these findings support a rich interpretation of preschoolers' trust behavior: children are not simply creating temporary mappings based on the situation, but instead making broader assumptions about the speaker (Scofield & Behrend, 2008). This rich interpretation fits well with both Harris' (2007) cognitive profile theory and with other selective trust research (Birch et al., 2007; Corriveau & Harris, 2009a; Fusaro & Harris, 2008; Poulin-Dubois & Chow, 2009).

Several other lines of research support the notion that preschool aged children's trust evaluations are both sophisticated and enduring attributions. Corriveau and Harris, (2009a) tested 3- and 4-year-old children at three separate time points (T1: at test, T2: 24 hours later, and T3: 1 week later) on a typical reliability paradigm (similar to Koenig et al., 2004). In line with previous research, children performed above chance on all three test questions (ask, endorse, and explicit judgment) during the initial testing session. Interestingly, children performed above chance levels at both T2 (1 day later) and T3 (1 week later). Therefore, both age groups not only identified and trusted the more accurate informant during the initial testing session, but also held this attribution in memory over an extended period of time.

In a follow-up study, Corriveau and Harris removed explicit judgment questions from the initial test period in an attempt to eliminate any external

reinforcement of speaker reliability. Even in this more cognitively demanding design, children in both age groups performed above chance at test and 4 days later, reliably choosing to trust the more accurate informant. Taken together, these findings suggest that children adjust their trust in others' based on minimal exposure to reliability information, and maintain these trustworthiness evaluations over time without external prompts. Moreover, these findings lend further support to the mentalistic attribution hypothesis (Harris, 2007), which argues that in order for children's attributions of reliability to be stable over time, they must consider the minimal accuracy information presented to be representative of the speaker's mental state, which is also stable over time (Corriveau & Harris, 2009a).

All of the research on preschoolers' selective trust capabilities discussed to this point has been built upon a design that presents children with a completely reliable speaker compared to either a completely unreliable speaker or ignorant speaker. While this design may be the least cognitively demanding and therefore the best at eliciting specific capabilities, it may lack ecological validity. It is rarely the case that in real world learning situations a speaker is either entirely reliable or entirely unreliable. Instead, children might encounter individuals whose level of reliability is not immediately apparent or has some variability, and consequently decisions about whom to trust may become difficult. Pasquini, Corriveau, Koenig, and Harris (2007) developed an experimental design to address this gap in the research.

In two experiments, Pasquini et al. (2007) expanded upon the typical reliability paradigm by presenting 3- and 4-year old children three unique accuracy to

inaccuracy ratios (100% correct vs. 25% correct, 75% correct vs. 0% correct and 75% correct vs. 25% correct) in addition to the standard 100% correct vs. 0% correct ratio. For instance, during familiarization trials (when speaker reliability was established) children were presented with four instances in which speakers either labeled familiar objects correctly or incorrectly. In the typical 100% vs. 0% condition, one speaker labeled all four objects correctly while the other labeled all four objects incorrectly. In the second condition (100% vs. 25%), one speaker was consistently correct across the four trials, whereas the other labeled only one out of four objects correctly. In the third condition (75% vs. 0%), one speaker provided correct labels on three out of four trials while the other was consistently inaccurate. Finally, in the fourth condition (75% vs. 25%), one actor labeled three out of four familiar objects correctly, whereas the other labeled one out of four correctly.

Selective trust was measured with three types of questions: ask, endorse, and explicit judgment. These questions were similar to those used in past reliability studies (Koenig et al*.,* 2004; Koenig & Harris, 2005; Scofield & Behrend, 2008). Children's performance on test trials was mediated by both age and performance on explicit judgment questions. Specifically, Pasquini and colleagues found that 3-year-olds displayed selective trust in two out of the four conditions: 100% vs. 0% and 100% vs. 25% (performing at chance in the other two conditions). Four-year-olds, however, demonstrated selective trust across all four conditions, choosing to ask and endorse the more reliable speaker regardless of the previous accuracy/inaccuracy ratio. Based on these findings, Pasquini and colleagues argued that 4-year-olds employ a refined statistical monitoring strategy that does not dismiss a speaker as

unreliable based on a single error. Instead, 4-year-old children are able to consider the number of mistakes committed by each informant and to compare the two in an attempt to assess the overall, relative reliability of both informants.

Children's sensitivity to a speaker's relative reliability, combined with their capability to reverse trust mappings and recall trustworthy attributions over time, suggests that preschoolers possess a sophisticated set of tools that they bring to learning situations. It is clear that children are not entirely credulous when evaluating testimony. If a given speaker provides unreliable testimony, children will track this information, encode the speaker's identity, make an attribution of untrustworthiness, and ultimately reduce the likelihood that they will seek out that unreliable informant during future learning opportunities. Children's reliability tracking and selective trust assessments, however, are not limited to information about reliable and unreliable output. Rather, children are keenly perceptive of various social cues that indicate the relative reliability of a speaker's claims.

Social cues. Preschoolers have demonstrated a powerful ability to track an individual's history of reliability and employ this information when making a trustworthy versus untrustworthy appraisal. In a typical real world scenario, however, children's attributions of others' credibility are, of course, not limited to information about a given speaker's reliable or unreliable output. Instead, children are both sensitive to and adept at using various social cues when assessing an individual's reliability. For example, when evaluating a speaker's assertive claims, preschool aged children are particularly sensitive to a speaker's cues of confidence (Jaswal, 2004; Jaswal & Malone, 2007). Jaswal and Malone (2007) demonstrated this sensitivity by

using a category induction procedure: an adult introduced an object by using a label that did not match the object's appearance (e.g., a hat-like object was labeled a "cup"). The adult's assertion was coupled with either cues of confidence or uncertainty. Uncertainty cues consisted of both verbal (e.g., the experimenter saying, "I think this is an X") and nonverbal signals (e.g., furrowed brow, hesitant voice). Three-year-olds were then asked to endorse either the function of the object as labeled by the experimenter, or the function of the object as it appeared.

Jaswal and Malone found that 3-year-olds were sensitive to a speaker's cues of confidence or uncertainty. Children who had witnessed an uncertain speaker provide a discordant label were less receptive to her unexpected label, choosing to endorse their own perceptual experience of the object and its function. In a follow-up study, Jaswal and Malone included additional conditions in which the speaker expressed distraction, ignorance, or both during the mismatched labeling. Three-year-olds exhibited credulity in both the distracted and ignorant conditions, whereas they demonstrated skepticism when a speaker displayed both distraction and expressed inattention as the reason for her error. These findings, taken together with earlier research on confidence cues (Jaswal, 2004), suggest that children monitor the level of confidence a speaker exhibits when making assertive claims, and use this social cue when deciding whether to trust that individual.

In addition to confidence cues, preschoolers are keenly perceptive of and sensitive to bystanders' nonverbal reactions to testimony (Fusaro & Harris, 2008). Fusaro and Harris found that 4-year-old children were more likely to endorse a speaker whose claims were accompanied by bystander cues of assent (e.g., nods and

smiles). Based on these bystander cues, children made trustworthiness evaluations even when the speaker never provided a correct label for a familiar object. Thus, even when a speaker has never provided a child with reliable information that he/she could check against personal experience, children's acute social awareness allows them to make an educated evaluation of a speaker as trustworthy or untrustworthy.

What other social cues might a child use to evaluate the trustworthiness of a given speaker? One interesting possibility is group membership. Only a few previous studies have addressed this question. For example, Jaswal and Neely (2006) tested 3- and 4-year-old children on a typical reliability paradigm, but included the comparison of an unreliable adult versus a reliable child. Results showed that, all things being equal, children trust adults first. However, when a child was obviously more accurate than an adult, children chose to trust their more reliable peer. This finding is important for two reasons. First, it demonstrates that children do consider group membership when making reliability judgments, showing a preference for adults based on their status as an adult. Second, it demonstrates that preschoolers are able to override their typical deference for a particular group when presented with relevant reliability information.

In perhaps the most direct measure of the potential influence of group membership on children's selective trust, Kinzler, Corriveau, and Harris (in press) asked whether children would rather learn from native-accented or foreign-accented speakers. Children were presented with films in which two female speakers read the first four sentences from a children's book. Each actress was a bilingual speaker of Spanish and English, and was recorded speaking English with an American accent

and English with a Spanish accent. This allowed the experimenters to control for numerous cues (e.g., visual appearance, voice quality, and comfort speaking). Once native- versus foreign-accent had been established, the two actresses silently demonstrated conflicting functions for a series of novel objects. Children's selective trust was defined by which object function they chose to endorse as the correct usage. Results showed that 4- and 5-year-olds' default decision was to trust the native-accented speaker (i.e., in-group member).

In an interesting follow-up experiment, Kinzler et al. removed any meaningful content from the accent exposure readings; thus, children could not use clarity of speech as a cue for selective trust. Even in this atypical scenario, where speakers read nonsense words in a native-accent, Kinzler and colleagues found similar results – children trusted the native-accented over the foreign-accented speaker. In sum, across two experiments, children were able and willing to use accent – a salient cue for an individual's group membership – as a guide when determining whom to trust. It is crucial to note that this study measured selective trust without including any history of accuracy/inaccuracy. Therefore, it appears that children will trust in-group members – defined by a shared native accent – in the absence of other sources of information (e.g., a history of reliability) that might potentially affect their appraisal.

Because of the sensitivity and efficiency of preschoolers' reliability tracking, it is sensible to suppose that children might flexibly alter their initial in-group preferences when given information about an individual's recent history of accuracy or inaccuracy. A recent study conducted by Corriveau and Harris (2009b) lends

support to this idea. Specifically, Corriveau and Harris investigated the relative weights of two cues during children's evaluations of trustworthiness: a form of social categorization (familiarity vs. unfamiliarity) and a recent history of reliability or unreliability. In two experiments, 3-, 4-, and 5-year-old children were tested on a typical reliability paradigm, but in a crucial manipulation, the two informants were a familiar teacher and an unfamiliar teacher. Initially, all age groups trusted the familiar teacher more, but only 4- and 5-year-olds were able to moderate that trust once presented with reliability information. In contrast, 3-year-olds continued to trust the more familiar teacher, even when she had proven less accurate during familiarization trials. Furthermore, Corriveau and Harris included an explanation probe, asking children why they thought a given speaker had inaccurately labeled a familiar object. Children's responses were markedly different depending on the social category of the speaker: they tended to attribute ignorance to the unfamiliar informant, but claimed that the familiar informant was merely pretending. These findings suggest that, even with the existence of an initial preference for trusting one social category over another, older preschoolers are able to effectively moderate their trust based on the presence of recent reliability information, thus exhibiting fairly flexible trust mappings. In addition, children appear to process previous reliability information differently based on an individual's social category.

The studies discussed earlier have shown that preschoolers are able to evaluate and use multiple types of information when making a trustworthiness evaluation (e.g., whether a speaker is a child vs. an adult, familiar vs. unfamiliar, reliable vs. unreliable, and/or a native vs. non-native speaker). What other types of

input might affect a child's decision about whom to trust? One viable prospect is in- or out-group membership. A wealth of research has demonstrated that children and adults readily categorize individuals into groups based on both meaningful characteristics (e.g., race, gender, language) and arbitrary divisions (e.g., red team vs. blue team). Furthermore, specific biases appear to result from these divisions. However, these lines of research have only just begun to address the question of whether group membership, beyond the categories of familiar vs. unfamiliar, adult vs. child and native vs. non-native speaker, affects a child's decision of whom to trust. Also, these attempts to address the influence of group membership have yet to pit prior reliability against in- or out- group status. Therefore, it is important to consider how children divide and classify the people they meet, and whether these categories might potentially outweigh a speaker's history of reliability, thus influencing one of the more important character judgments a person can make – whether to trust or distrust another individual.

Social Categorization and Minimal Group Bias

Social Categorization

Natural Groups. A propensity to categorize individuals based on some meaningful characteristic has been proposed as a potential addition to the set of core knowledge systems. Researchers suggest that humans enter the world with a natural tendency to divide others into social categories and along in- versus out-group lines (Spelke & Kinzler, 2007). Furthermore, evolutionary psychologists have argued that humans evolved with a biological system for tracking coalitions and alliances

(Cosmides, Tooby, & Kurzban, 2003). Do children demonstrate these categorization behaviors, and if so, what effects might these apparently natural and biological systems have on children's interpretations of the physical and social world?

Recent developmental research indicates that young children do in fact differentiate between individuals based on characteristics such as race, gender, age, and language. In the case of gender, evidence from infant research demonstrates that 3- to 4-month-olds make a distinction between different gender faces and show a preference for the gender of their primary caregiver (Quinn, Ayahr, Kuhn, Slater, & Pascalis, 2002). Later in development, children begin to demonstrate a marked preference for their own gender when making decisions about with whom to play. Maccoby and Jacklin (1987) found that 4.5-year-old nursery school children spent three times as much time playing with same-sex playmates as they did with cross-sex partners. Finally, by 5-6 years of age, children explicitly judge their gender more positively and apply rigid gender category norms (Ruble & Martin, 1998 as cited in Cameron, Alvarez, Ruble, & Fulgini, 2001). Taken together, these findings suggest that gender serves as a salient marker, allowing children to divide the social world into two groups, and when these groups are clearly defined, children reliably prefer their own gender's in-group.

Convergent evidence of young children's aptitude for categorization can be found in research on native language preferences. Some researchers have argued that language may be a stronger cue than race to social group membership, as contact with different racial groups were not as likely when these categorical processing systems evolved (Cosmides, Tooby, & Kurzban, 2003). To address this hypothesis, Kinzler &

Spelke (2005) presented 6-month-old infants with films of two women's faces who were both fluent in English and Spanish. After the women spoke in alteration, infants were presented with images of the two women side by side. The dependent measure was infants' looking behavior, and results showed that infants preferred to look at the woman who spoke their native language (English).

In a follow-up study, 5-year-old children demonstrated a similar preference for native language speakers when making friendship decisions. Children also favored speakers with native accents as opposed to foreign accents, and, interestingly, children chose to be friends with other-race speakers who spoke with a native accent over own-race speakers with a foreign accent (Kinzler, Shutts, DeJesus, & Spelke, 2009). Overall, children's acute sensitivity to the language/accent of a speaker, coupled with their strong native language/accent preference, suggests that it is an additional, viable marker of group membership similar to gender.

A preference for own-race-faces, similar to infants' preference for faces matching the gender of their primary caregiver, can be found as early as 3 months of age in racially homogenous populations (Kelly, Quinn, Slater, Lee, Gibson, Smith, Liezhong, & Pascalis 2005; Bar-Haim, Ziv, Lamy, & Hodes, 2006). Thus, at an early age infants are able not only to distinguish between racial groups, but additionally to demonstrate a marked looking preference for their own race's in-group.

Support for young children's early racial categorization and bias is not limited to behavioral evidence from looking preferences. Using a modified version of the Implicit Association Test (IAT), Baron & Banaji (2006) demonstrated that 6- and 10-year-old Caucasian children, as well as adults, display implicit pro-White and anti-

Black biases. Baron and Banaji measured the relative strength of association between the target concept of race and an attribute concept such as words with good or bad meanings. Caucasian participants displayed a stronger association between members of the Black racial group and negative words. Interestingly, the relevant amount of explicit bias decreased with the age of the participants, but implicit attitudes did not vary. This suggests that implicit biases are formed early in life and are stable over time, and what adjusts is children's understanding of social norms related to self-presentation (Baron & Banaji, 2006).

In sum, the findings associated with children's categorization of individuals based on gender, language, and race make a strong case that children organize the social world through the use of categories, and consequently form distinct in-group preferences based on these categories.

Although a substantial body of work provides evidence of children's inclination to categorize and from in-group preferences, it is still unclear how children may use social categories when thinking about similar or dissimilar others. In an effort to address this question, Diesendruck and haLevi (2006) asked both adults and children to make inferences about a novel character's psychological properties based on either similar social category membership (e.g., gender, ethnicity) or similar personality traits (e.g., niceness, shyness). Children were presented with a triad of characters: one was the target; one was the social category match; the third was the personality match. Results showed that children chose to make psychological inferences (e.g., assessing a character's desire to play a certain game) based on social category membership. Interestingly, for the population studied (Israeli 5-year-olds),

the social categories of "Jew" and "Arab" were the most inductively powerful. This suggests that young children not only categorize the social world, but also are likely to use that information to make quick decisions about the psychological properties of group members.

In addition to supporting a general in-group preference, social categorization results in measurable cognitive, social, and affective consequences. These consequences include an increased perception of between-group differences and within-group similarity (Bigler, Jones, & Lobliner, 1997; Rothbart, Davis-Stitt, & Hill, 1997), an increased perception of out-group homogeneity (Park & Rothbart, 1982), and biased memory processes (Meissner & Brigham, 2001). Furthermore, group members may commit "the fundamental attribution error" when evaluating out-group behavior, disproportionately attributing negative out-group actions to their dispositions as opposed to the situation (Hamilton & Trolier, 1986).

The impact of group membership, however, is not limited to specific cognitive processes and attributions. Rather, at a young age, children's friendship decisions are influenced by others' social categories: a series of previous studies indicates that children display an awareness of physical characteristics that define racial groups at an early age and that they are willing to apply that information (Aboud, 1988; Kowalski & Lo, 2001). For instance, Kowalski and Lo (2001) found that 3- to 5-year-old Taiwanese children chose Asian children – represented by photographs in an array containing photographs of White, Black, and Asian children – significantly more often than chance would predict when asked to choose a playmate. Therefore, children as young as 3 years of age were able to functionally distinguish between

racial groups and use this information to influence friendship choices, choosing to be friends with in-group members more often. In sum, the division of the world into social categories influences (a) psychological inferences about group members, (b) specific cognitive processes (e.g., memory, attributions), and (c) more general processes and decision making (e.g., friendship choices).

Minimal Groups. The impact of biases engendered by in-group/out-group categorization, however, is not limited to pre-existing groups. A wealth of research conducted with adults employing the Minimal Group Paradigm (MGP), which places individuals into different categories based on arbitrary and seemingly meaningless characteristics (e.g., shirt color), has demonstrated the surprisingly minimal requirements necessary to create intergroup bias. Participants in these studies, though often lacking any conflict of interests or previously existing hostility and categorized randomly, nonetheless exhibit strong in-group preferences marked by: (a) in-group favoritism, (b) out-group discrimination, and (c) competitive treatment of the out-group (see Brewer, 1979 for a review).

A recent demonstration of children's susceptibility to intergroup bias based on the MGP comes from developmental research conducted by Bigler and colleagues. In a series of studies, children were divided into two novel groups differentiated by shirt color, and then spent a predetermined amount of time (usually between 3 to 6 weeks) interacting with both in- and out-group members in a classroom setting (Bigler et al., 1997; Bigler, Spears Brown, & Markell, 2001; Patterson & Bigler, 2006). At the end of the allotted time, children were asked several series of questions to address any biases that may have arisen. Variables such as how frequently the two groups were

acknowledged or referenced and whether the groups were based on random or biological characteristics (e.g., hair color) were investigated. Across several studies, children were quick to develop intergroup biases even when the groups were randomly assigned. For instance, children were inclined to attribute more positive traits to *all* members of their in-group, unwilling to assign negative traits to *any* members of their in-group, and predicted that their in-group would perform better in a series of three contests against the out-group. Thus, even when arbitrarily categorized, children demonstrated a distinct, enhanced image of their in-group that extended to concrete predictions of performance on an unrelated task.

In sum, humans readily categorize others into both natural and minimal groups, and, consequently, future interactions with categorized individuals are influenced by their status as either a similar in-group member or a dissimilar out-group member. Furthermore, the MGP has proved sufficient to elicit similar intergroup biases that result from natural group distinctions. With regard to trust, research in this area has demonstrated that intergroup biases most reliably influence a common set of dimensions, which include trustworthiness, honesty, and loyalty. For instance, Dion (1973) had participants evaluate both their in-group and out-group on a series of dimensions (e.g., attractiveness, pleasantness, trustworthiness) after participating in an intergroup Prisoner's Dilemma Game. Dion found that intergroup bias in favor of the in-group was greatest on the dimension of trust. Based on these findings, it seems that social categorization along in- versus out- group lines is quite capable of altering trustworthiness evaluations. Therefore, this thesis will employ the MGP to address this possibility and, specifically, to test the impact of in- and out-

group status on young children's well-documented capability to selectively trust more accurate individuals. Will group membership in the MGP alter children's ability to monitor a speaker's accuracy vs. inaccuracy and subsequently bestow trust efficiently? Will in-group favoritism prove too powerful a cue, thus requiring overwhelming evidence of out-group reliability in order to establish trust? Or, will a previous history of accuracy take priority in children's evaluations of another's trustworthiness?

An Experimental Investigation of the Influence of Group Membership on Preschoolers' Selective Trust

Young children's ability to track and use a speaker's history of accuracy/inaccuracy when determining trustworthiness has been well documented. By the age of 4, children are able to employ a reliability tracking system that is capable of: (a) differentiating between a reliable and unreliable speaker even when the difference between the two is slightly ambiguous (Pasquini et al., 2007), (b) reversing trust when a speaker is later determined to be unreliable (Scofield & Behrend, 2008), and (c) maintaining trustworthy versus untrustworthy attributions over time (Corriveau & Harris, 2009a). In addition, children are able to attend to various social cues such as bystanders' verbal/nonverbal signals of assent and dissent (Fusaro & Harris, 2008) and speakers' confidence cues (Jaswal, 2004; Jaswal & Neely, 2006) when evaluating testimony. Taken together, these findings suggest that humans develop a relatively powerful and reliable system to evaluate the trustworthiness of others at a young age. But how precise and resilient is this system? Do other categorization processes influence or even take priority over young children's selective trust in more reliable speakers?

One feasible prospect is humans' natural inclination for social categorization. A wealth of research from numerous fields has demonstrated that adults and children will readily categorize the social world based on gender, race, age, and language, and will readily develop in-group preferences based on these group distinctions (Kinzler, Shutts, & Correll, in press). Adults and children do not merely divide others into similar groups and prefer their in-groups; moreover, social processing is subsequently

influenced by these groupings. Specifically, social categorization results in an increased perception of between-group differences and within-group similarity (Bigler, Jones, & Lobliner, 1997; Rothbart, Davis-Stitt, & Hill, 1997), an increased perception of out-group homogeneity (Park & Rothbart, 1982), and biased memory processes (Meissner & Brigham, 2001). Finally, social categorization can lead individuals to commit "the fundamental attribution error" when evaluating others' actions, disproportionately attributing negative out-group behaviors to their dispositions as opposed to the situation (Hamilton & Trolier, 1986).

The intersection of humans' seemingly ubiquitous system of social categorization and children's sophisticated ability to selectively trust leads to two important questions. Are the two systems mutually exclusive? If not, which system takes priority when the two interact or conflict? Typically, this interaction has been addressed by economic and management research conducted with adults that use prisoner dilemma and allocation games to measure trust of in- versus out-group members, which has shown mixed results (Buchan et al., 2006; Brewer & Silver, 1978; Dion, 1973; Guth et al., 2007). In the developmental literature, researchers have begun to investigate the priority of group membership processes during trust evaluations. In a recent experiment, Kinzler et al. (in press) assessed whether children would default to trusting a native accented-speaker over a foreign-accented speaker when deciding whom to learn from. Across two studies, children chose to trust and learn from a native-accented speaker, thus demonstrating a marked in-group preference during trust evaluations (with no conflicting information provided). This study did not include any previous history of accuracy or inaccuracy, which might

potentially outweigh any initial in-group trusting preference. In another recent study, Corriveau and Harris (2009b) developed an experimental design to address this possibility. Children were asked to decide between bestowing trust on a familiar or an unfamiliar teacher, who had either demonstrated previous accuracy or inaccuracy. Initially, in the absence of prior reliability information, young children chose to trust the more familiar informant. Once given evidence of prior accuracy or inaccuracy, 4- and 5-year-olds (but not 3-year-olds) were able to accurately alter their trust, directing questions towards and endorsing the more accurate informant regardless of their familiar vs. unfamiliar status. In sum, children demonstrated flexible selective trust systems, capable of overriding initial preferences for trusting a familiar informant when provided with pertinent prior reliability information. Although this study addressed the relative importance of one distinct form of social categorization during trustworthiness evaluations, it did not assess whether shared in-group status or conflicting out-group status influences children's selective trust. Thus, the direct interaction of prior reliability and in- and out-group membership has not yet been investigated.

Here, I directly address this interaction with two experiments. In Experiment 1, children were asked to decide between bestowing trust on a reliable in-group member or an unreliable in-group member, across a variety of accuracy/inaccuracy ratios, in order to assess any potential influence of shared within-group membership on children's selective trust. To create the two distinct groups, I employed the Minimal Group Paradigm, which has been shown to elicit intergroup bias comparable to that elicited by pre-existing groups (Brewer, 1979). If reliability tracking and

selective trust operates separately from social categorization, then children should have no trouble endorsing and learning from the more reliable speaker. However, if within-group membership influences trust processing, then children might demonstrate less sensitivity to the available reliability information, trusting either informant because of their shared in-group status. In Experiment 2, children were asked to decide between bestowing trust on a reliable out-group member or an unreliable in-group member. As in Experiment 1, if prior reliability information is more salient, then children should succeed in selectively trusting the more reliable speaker despite their out-group status. However, if social categorization and intergroup biases are more salient, then children should demonstrate less sensitivity to the reliable versus unreliable distinction, possibly trusting an unreliable in-group member over a reliable out-group member.

Experiment 1

Experiment 1 was designed to examine the influence of shared in-group status on young children's selective trust in relatively more accurate informants. To assess the comparative weight of these two forms of categorization (in-/out-group vs. reliable/unreliable), children were presented with two minimal groups (the Red Group vs. the Blue Group), but, crucially, children were only placed into the Blue Group. Children were then shown films that depicted a reliable in-group informant pitted against an unreliable in-group informant across four different accuracy/inaccuracy ratios to see who children would rather direct questions towards and subsequently learn from (i.e., who they would trust). The first condition resembled a typical reliability paradigm: across four accuracy trials one speaker was consistently accurate

when naming familiar objects (100% correct), while the other speaker was consistently inaccurate (0% correct). The remaining three conditions consisted of the following accuracy/inaccuracy ratios: 100% correct vs. 25% correct, 75% correct vs. 0% correct, and finally 75% correct vs. 25% correct. These ratios were taken from Pasquini et al. (2007), and used to address whether increasing the ambiguity of a speaker's history of reliability influenced children's trustworthiness evaluations of in-group members.

To explore whether 4-year-olds' selective trust was altered by the inclusion of shared in-group membership, children were presented with a series of four test trials that included *Ask* and *Endorse* questions. Ask trials prompted children to select which informant they would rather ask for further information about a novel object. Endorse trials asked children to indicate which informants' novel label for the unfamiliar object they agreed with. In addition, two sets of *Explicit Judgment* (EJ) probes were asked at two distinct times during the experiment: the first set immediately following the presentation of prior reliability information and the second set at the very end of the task. EJ trials assessed which informant children thought was "better" at naming both the familiar and the novel objects, and provided a measure of how children evaluated the speakers at different points of the experiment. If children display greater sensitivity to monitoring the relative reliability of each speaker, I expect 4-year-olds to selectively trust (i.e., ask and endorse) the more accurate speaker, and explicitly judge the more accurate speaker as being more reliable across all four conditions, regardless of the speakers' in-group statuses. If shared group membership and consequently in-group bias alters selective trust capabilities, then children might

demonstrate less sensitivity to prior reliability information, bestowing trust on both informants based on their status as in-group members.

Method

Participant information: Twenty-two 4-year-old children were tested in this Experiment (M = 4 years 4 months; range = 3 years 9 months to 5 years 2 months; 13 girls). Children, for the most part, were recruited from the local Middletown, CT area and tested in a quiet laboratory setting at Wesleyan University. Several children were recruited and tested at local Preschools in the Middletown area. The majority of children were White.

Establishing Group Membership: At the start of testing, each child was informed of the two different groups that were involved in the game - the Blue Group and the Red Group. Children were then told that as a special part of the game they would be placed into the Blue Group. Each child was placed into the Blue group to ensure that they shared in-group status with the two actors in the films. To increase the salience of their group status, children were asked to wear both a solid blue jersey and solid blue wristbands.

Children were then asked a series of questions aimed at reinforcing their group membership. The experimenter presented the child with a sequence of four photographs containing both members of the in-group (individuals wearing blue shirts) and members of the out-group (individuals wearing red shirts), and asked the child, "Is she/he in your group or the other group?"

Selective Trust Task: Once group membership had been effectively established, children proceeded to the selective trust task, which was adapted from Pasquini et al. (2007). Each child was tested across four conditions, with the order of the conditions counterbalanced across participants.

Children were placed in front of a laptop with a still frame of two actors, both wearing solid blue t-shirts (both actors were members of the child's in-group) and seated to the left or right of an object sitting on a small wooden table. Next, the experimenter pointed to the actor on the right or the left and asked, "Is she/he in your group or the other group?" The same procedure was repeated for the other actor on the screen. Then the experimenter introduced the task and said,

> *I've got these two friends. See? They both have blue shirts, just like you. They're going to show you some things and tell you what they are called. I want you to listen very carefully and then I'm going to ask you some questions. Let's watch.*

Each child saw four separate films, which corresponded with the four separate conditions. Each film consisted of eight clips respectively: four *accuracy* trials in which the objects were familiar (e.g., ball, phone) and four *test* trials with novel/unfamiliar objects (e.g., arm sling, yellow toy column; see tables 1 and 2 for full lists of familiar and novel objects used). At the beginning of every clip, children were presented with an enlarged still frame of the object in order to ensure clear visual recognition of the object during the clip. The order of trials within a given film was maintained throughout participants (Tables 1 and 2).

--------- TABLES 1 AND 2 ABOUT HERE ---------

Within a given a film/condition the actors remained constant, but across conditions the actors varied. Additionally, two separate sets of gender-specific videos

were made with all male and all female actors respectively. This was to ensure that the gender of the actors in each film corresponded to the gender of the participant. Thus, each female participant saw only female actresses, while each male participant saw only male actors. In each film, the actors and actresses were similar in age, ethnicity, and appearance.

Accuracy trials. Each of the four accuracy trials began with an enlarged still frame of the familiar object on the laptop screen. The experimenter then said, "Hmm, a new object. Let's see what they say about that." Then children watched a clip with two actors and the familiar object located on a table between them. The first actor turned to the familiar object and while maintaining visual contact with the object said, "That's a _____." Then the second actor, who responded with a different label, turned to the object and said, "That's a _____." Within each film, the actor who spoke first (left v. right) alternated across the four *accuracy* trials; between subjects, the actor who spoke first (left v. right) for a given condition was counterbalanced. At the end of the four *accuracy* trials, children were asked a series of *explicit judgment* questions. The experimenter said,

i. *"Was she/he (*pointed/referenced actor on computer screen) *good at naming the objects or was she/he not very good at naming the objects?"* (Actor 1)
ii. *"Was she/he good at naming the objects or was she/he not very good at naming the objects?"* (Actor 2)
iii. *"Who was better at naming the objects: him/her (*pointed) *or him/her* (pointed)*?"*

The order of which actor was referenced first was counterbalanced across participants. Familiar objects used in the experiment were taken from Pasquini et al. (2007), which matched object names for age of acquisition (Fenson et al., 1994).

There were four separate conditions; therefore, four distinct films were created. In Film 1 (100% v. 0% condition), one actor named all four familiar objects during the accuracy trials correctly (100% correct), while the other named all four familiar objects incorrectly (0% correct). After each trial, the experimenter paused the video and said, "The girl/boy on the right called it a _____, and the girl/boy on the left called it a _____. What do you think it is called?" The experimenter pointed to the screen to make it clear which actor provided which label for the familiar object. Once the child provided their own label, they were asked to point to whom they thought labeled the object correctly. For half the participants, the actor on the left was 100% reliable. For the other half, the actor on the right was 100% reliable.

Film 2 (100% v. 25% condition) was identical in structure to Film 1 except one actor named all four familiar objects correctly (100% correct), whereas the other actor named one out of the four familiar objects correctly (25% correct). On the trial in which both actors overlapped and provided accurate labels, each actor provided the same accurate label. Following each video clip and at the end of the four *accuracy* trials, the experimenter asked the same questions as in Film 1. The particular trial in which both actors correctly named the familiar object varied randomly across participants.

In Film 3 (75% v. 0% condition), one actor named three out of the four objects correctly (75% correct), while the other actor named all four objects incorrectly (0% correct). On the trial in which both actors provided inaccurate labels, they provided different inaccurate labels. Following the video clips, the experimenter asked the same questions as in Films 1 and 2. For half the participants the actor on the

right was 75% correct, and for half the participants the actor on the left was 75% correct. The trial in which both actors provided inaccurate labels varied randomly across participants.

For Film 4 (75% v. 25% condition), one actor named three out of four objects correctly (75% correct), and one actor named one out of four objects correctly (25% correct). In this condition, the actors' correct or incorrect labels never overlapped (i.e., there were no clips in which both actors made errors or both were correct). The same questions were asked as in Films 1, 2, and 3. The position of the clip in which the mostly accurate (75% correct) actor made an error and the mostly inaccurate (25% correct) actor provided the correct label was varied randomly across participants.

Test Trials. After the four *accuracy* trials, children were shown four successive *test* trial clips. Each of the four test trials began with a still frame of the unfamiliar object on the laptop screen, and the experimenter asked the child, "Do you know what that is called?" Children were given an opportunity to reply and then were asked, "I bet one of these people knows what it is called. Which one would you like to ask?" In the highly unusual case in which children provided their own label for the unfamiliar object, they were told, "Actually, I don't think that's exactly what it is called. I bet one of these people knows what it is called. Which one would you like to ask?" After this question, children viewed a short video clip in which the first actor turned his/her gaze towards the unfamiliar object on the table and provided a novel label (e.g., merval). Then, the second actor turned to the object and provided a distinct novel label (e.g., zav; for a full list of unfamiliar objects and novel labels see

Table 2). The order of the test trials, as shown in Table 2, remained constant across participants in all four conditions.

After each video clip (one test trial), the experimenter paused the film and children were asked, "She/he (experimenter points to actor) called it a _____, and she/he (experimenter points to the other actor) called it a _____. What do you think it's called, a _____ or a _____?" At the end of the four test trials, the explicit judgment questions were asked (as in the accuracy trials).

Finally, at the end of the experiment, children were asked, "Which group do you belong to?" This was asked to ensure that the children maintained an understanding of their own group membership both at the beginning at the end of the task.

Results

I first report on children's performance on the two separate group membership assessments. Next, children's performance during accuracy trials is discussed. I then analyze children's responses to the three test questions (explicit judgment, ask, and endorse) and compare the results to chance performance. Finally, children's responses to the two sets of explicit judgment questions are examined separately, in order to address any differences between children's explicit assessment of speaker reliability immediately following the presentation of accuracy information and at the end of the task.

Performance on group membership assessments. At the beginning of the task, all children readily accepted their group membership (Blue Group), and accurately

identified which group they belonged to. In addition, all children were able to correctly identify and label novel individuals as being members of "their" group or the "other" group. Finally, each participant accurately identified which group they belonged to when asked at the very end of the task.

Performance during accuracy trials. Across all four conditions, the majority of children were able to correctly recognize and produce the accurate label for all the familiar objects. This was the label provided by the more accurate speaker. However, one participant chose to endorse the label provided by the inaccurate speaker on all four conditions. This child was excluded from any further analysis. In addition, 2 children erred during the 75% correct versus 0% correct condition when both speakers provided inaccurate labels for the familiar object. For the most part children chose to endorse the informant who had demonstrated previous accuracy. These children were included in the sample.

Comparisons to chance for the three test questions. The results of children's performance on the three test questions are presented in Table 3. Proportion correct and comparisons to chance for the Explicit Judgment (EJ), Ask, and Endorse questions were calculated. EJ scores were calculated based on the final series of questions, asked at the end of each condition. Children performed above chance levels on the *EJ* questions in the 100% versus 0% condition, and the 100% versus 25% condition. However, they performed at chance levels in the 75% versus 25% condition and the 75% versus 0% condition. For both Ask and Endorse questions children responded accurately and above chance levels in the 100% versus 0% correct condition and the 75% versus 0% correct condition. However, performance did not

significantly differ from chance in the 100% versus 25% correct condition and the 75% versus 0% correct condition.

--------- TABLE 3 ABOUT HERE ---------

Performance on both sets of explicit judgment questions. The results of children's responses to the first series of EJ questions, asked immediately following the 4 familiarization trials, are presented in Table 4. A mean score (range 0-1) for each of the three EJ questions was calculated. Immediately following the familiarization trials, children performed above chance on all three EJ questions across all four conditions, correctly identifying which speaker was more accurate and "better" at labeling the familiar objects.

The results for the second series of EJ probes, asked at the end of the task, are displayed in Table 5. Children performed above chance in the 100% versus 0% correct condition, the 100% versus 25% correct condition, and on the first question ("Was she/he good or not very good? – Referring to the more accurate speaker) in the 75% versus 25% correct condition. However, children were at chance on the final 2 EJ questions in the 75% versus 25% correct condition and on all three questions in the 75% versus 0% correct condition.

--------- TABLES 4 AND 5 ABOUT HERE ---------

Discussion

The main goal of Experiment 1 was to investigate how children evaluate and use the relative prior reliability of two in-group members. Will shared in-group status alter how children process prior reliability information when making trustworthiness

evaluations? These results support three findings of interest. First, 4-year-olds displayed a keen ability to identify and categorize themselves and others into the novel Blue and Red groups. Second, similar to previous selective trust research (Birch et al., 2007; Koenig & Harris, 2004; Koenig, Clement, & Harris, 2004; Pasquini, Corriveau, Koenig, & Harris, 2007), 4-year-old children were generally accurate when identifying and tracking a more versus less reliable informant, thus demonstrating a relatively sensitive reliability tracking system. Finally, 4-year-olds' ability to use a speaker's previous history of accuracy/inaccuracy, when deciding whether to later bestow trust, was desensitized by the inclusion of in-group membership.

Across all participants, children had little difficulty accepting their group membership and accurately differentiating similar in-group versus dissimilar out-group members. Furthermore, children's ability to recall their group affiliation when asked at the end of the task demonstrates, at the very least, that children maintained this in- versus out-group distinction throughout the experiment.

Children's performance on the two sets of Explicit Judgment (EJ) questions suggests an interesting effect of the group membership manipulation on children's reliability tracking. First, consistent with previous research, 4-year-olds had little difficulty tracking and distinguishing between a more versus a less reliable speaker, as evidenced by their successful performance on the first set of EJ questions (posed immediately following the presentation of prior reliability information). However, in the second set of EJ questions, presented 5-8 minutes after children witnessed the relevant prior reliability information, children demonstrated less accuracy when

determining whether an informant had been previously reliable or unreliable. Initially, this result may seem logical – when farther removed from the relevant information children should demonstrate less accurate recall. However, numerous studies have found that children typically show little difference in their patterns of performance on separate sets of EJ questions (Fusaro & Harris, 2008; Koenig et al., 2004; Koenig & Harris, 2005 Pasquini et al., 2007), usually choosing to collapse performance on both sets of questions into one EJ score. Therefore, based on children's performance when evaluating two in-group members (Exp. 1), it may be that 4-year-olds' reliability tracking capabilities were desensitized by the inclusion of group membership.

For an even stronger demonstration of shared in-group membership's influence, I next turn to children's performance on the Ask and Endorse probes (i.e., whether children displayed selective trust). The results, which are displayed in Table 3, for both the 100% versus 0% correct and 75% versus 0% correct conditions are in line with Pasquini et al.'s (2007) findings that 4-year-olds are able to successfully track and use an informant's prior history of accuracy to selectively direct questions towards and learn from a more reliable informant. Children's inability to demonstrate selective trust in the 100% versus 25% correct and the 75% versus 25% correct conditions, however, suggests that in this study children were unable to display an equivalent sensitivity when employing their selective trust systems, choosing to ask and endorse both the more/less reliable informants at a rate that did not differ from chance. Once again, it is crucial to note that without the addition of group membership, 4-year-olds were able to effectively trust a relatively more reliable speaker, across all four accuracy/inaccuracy ratios (Pasquini et al., 2007).

In sum, these results suggest that 4-year-olds are capable of monitoring and using prior reliability information when deciding whom to trust from within their in-group during unambiguous situations. Interestingly, however, children were unable to track and use similar information provided by in-group members when the evidence of a speaker's previous reliability was not overwhelming. It appears that when differentiating between the trustworthiness of two in-group speakers, 4-year-olds require significant separation between a reliable versus an unreliable informant in order override their in-group status and categorize the two as trustworthy or untrustworthy, respectively. While Experiment 1 suggests that in-group status desensitizes children's selective trust systems, it is unclear how children evaluate and weigh the prior reliability of out-group members. Which cue will prove more salient – prior reliability or in-group status? Is a relatively reliable out-group member more trustworthy than a relatively unreliable in-group member?

Experiment 2

The findings from Experiment 1 suggest that the inclusion of group membership information, even if the in-group is not directly compared to an out-group, nonetheless influences children's selective trust capabilities. Four-year-olds in Experiment 1 demonstrated less sensitivity to prior reliability information than they did in a previous study that did not include group membership (Pasquini et al., 2007), when evaluating the trustworthiness of two in-group members. The question of how children might evaluate and compare the prior reliability of in-group and out-group members, however, still remains.

Experiment 2 was designed to investigate this question. The main goal was to assess whether children would trust a more reliable out-group member over a less reliable in-group member. Will group membership prove too powerful a cue for prior reliability to overcome? Or, will minimal groups provide insufficient evidence for children to dismiss a more reliable speaker? To address this scenario, children were placed into one of two minimal groups (the Blue Group or the Red Group) and presented with films, similar to those used in Experiment 1, which established the relative reliability of two speakers. Crucially, the more reliable informant was always a member of the child's *out-group*. As in Experiment 1, the following 4 accuracy/inaccuracy ratios were used to address any influence of ambiguous prior reliability information: 100% vs. 0% correct, 100% vs. 25% correct, 75% vs. 0% correct, and 75% vs. 25% correct.

Children's selective trust was probed with the same Ask and Endorse questions, posed during the four test trials. In addition, the same two sets of Explicit Judgment (EJ) questions were asked to gain an understanding of which informant the children thought was "better" at naming the objects. Choosing to use two sets of EJ questions, temporally removed from one another and from the initial presentation of reliability information, provided insight into how children's evaluations of the speakers changed throughout the experiment. I might expect that in the unambiguous condition (100% correct vs. 0% correct) 4-year-olds' selective trust would be unaffected by in- or out-group status. Furthermore, it is sensible to suppose that the desensitization of children's selective trust displayed in Experiment 1 might emerge

even more strongly in the more ambiguous conditions (75% correct vs. 25% correct and 100% correct vs. 25% correct).

Method

Participant Information: Twenty 4-year-old children were tested in this experiment (M = 4 years 4 months; range = 4 years 0 months to 4 years 11 months; 13 girls). Children, for the most part, were recruited from the local Middletown, CT area and tested in a quiet laboratory setting at Wesleyan University. Several children were recruited and tested at local Preschools in the Middletown area. The majority of children were White.

Establishing Group Membership: As in Experiment 1, at the start of testing, each child was informed of the two different groups that were involved in the game - the Blue Group and the Red Group. Children were then told that as a special part of the game they would be placed into either the Blue Group or the Red Group. In contrast to Experiment 1, however, children were placed into both the Blue and Red Group (as opposed to only the Blue Group). Children's group membership was counterbalanced so that half of the sample became members of the Red Group while the other half became members of the Blue Group. To increase the salience of their group status, children were again asked to wear both a solid blue/red jersey and solid blue/red wristbands.

Children were then given a series of questions aimed at reinforcing their group membership. Questions were identical to those used in Experiment 1. The experimenter presented the child with a sequence of four photographs containing both members of the in-group (individuals wearing blue shirts) and members of the out-

group (individuals wearing red shirts), and asked the child, "Is she/he in your group or the other group?"

Selective Trust Task: Once group membership had been effectively established, children proceeded to the selective trust task, which was similar to the overall design of Experiment 1, but included members of both the in- and out-group (i.e., blue and red shirted speakers). Each child was tested in all four conditions, with the order of the conditions counterbalanced across participants.

Children were placed in front of a laptop with a still frame of two actors, one wearing a solid blue t-shirt and the other wearing a solid red t-shirt (one actor was always a member of the child's in-group while the other was always a member of the child's out-group). The two actors were seated to the left or right of an object sitting on a small wooden table. Next, the experimenter pointed to the actor on the right or the left and asked, "Is she/he in your group or the other group?" The same procedure was repeated for the other actor on the screen. Then the experimenter introduced the task and said,

> *I've got these two friends. See? One has a blue/red shirt just like you, and the other has a blue/red shirt not like you. They're going to show you some things and tell you what they are called. I want you to listen very carefully and then I'm going to ask you some questions. Let's watch.*

The stimuli and procedure used in Experiment 2 were nearly identical to those used in Experiment 1. Again, each child saw four separate films comprised of eight clips respectively. Each film consisted of four *accuracy* trials where the objects were familiar (e.g., ball, phone) and four *test* trials where the objects were novel/unfamiliar (e.g., arm sling, yellow toy column; see Tables 1 and 2 for full lists of familiar and unfamiliar objects used). At the beginning of every clip, children were presented with

an enlarged still frame of the object in order to ensure clear visual recognition of the object during the clip. The order of trials within a given film was maintained throughout participants (Tables 1 and 2).

The same actors were used in Experiment 2 as those used in Experiment 1. Within a given a film/condition the actors remained constant, but across conditions the actors varied. Once again, to control for the child's gender, two separate sets of videos were made with all male and all female actors respectively; thus, each female participant saw only female actresses, while each male participant saw only male actors. In addition to the gender-specific videos, four additional sets or types of videos were made, which controlled for the actors' positions on the screen and shirt color.

While the position of the actors within a given condition/video did not vary, two sets of videos were created: one set in which the actor on the left wore the red shirt (actor on the right wore the blue shirt) and another set in which the actor on the right wore the red shirt (actor on the left wore blue shirt). In addition, within these two sets of videos, the accurate speaker varied systematically between the two groups. Thus, there were four distinct types of videos that a child could see depending on their group membership (established at the beginning of the task): (a) accurate Red Group speaker on the left, (b) accurate Red Group speaker on the right, (c) accurate Blue Group speaker on the left, and (d) accurate Blue Group speaker on the right. These four types of videos varied systematically across participants. For instance, if a child was a Blue Group member, he/she would only view films in which a Red Group speaker was more reliable. Further, each child saw four films corresponding to the

four conditions (i.e., the four accuracy/inaccuracy ratios), and, within those four films, the side of the screen in which the more reliable speaker appeared was counterbalanced (i.e., for two films the reliable Red Group speaker would appear on the left and for two films the reliable Red Group speaker would appear on the right). This was done to avoid any potential side of the screen bias.

Accuracy trials. The design and procedure of the accuracy trials was similar to Experiment 1, except that, as a key manipulation, in all four conditions the more reliable speaker was always a member of the child's out-group. Each accuracy trial began with an enlarged still frame of the familiar object on the laptop screen. The experimenter then said, "Hmm, a new object. Let's see what they say about that." Then children watched a clip with two actors and the familiar object located on a table between them. The first actor turned to the familiar object and while maintaining visual contact with the object said, "That's a _____." Then the second actor, who responded with a different label, turned to the object and said, "That's a _____." In each film, the order of which actor spoke first (left v. right) alternated across the four accuracy trials. At the end of the four accuracy trials, children were asked a series of *explicit judgment* questions that were slightly modified from Experiment 1. The experimenter said,

iv. *"Was the girl/boy in the red shirt (*pointed/referenced actor on computer screen) *good at naming the objects or was she/he not very good at naming the objects?"* (Actor 1)
v. *"Was the girl/boy in the blue shirt good at naming the objects or was she/he not very good at naming the objects?"* (Actor 2)
vi. *"Who was better at naming the objects: her/him (*pointed) *or her/him* (pointed)*?"*

The same familiar objects were used in Experiment 2 as those used in Experiment 1.

There were four separate conditions; therefore, four distinct films were created. In Film 1 (100% v. 0% condition), one actor named all four familiar objects during the accuracy trials correctly (100% correct), while the other named all four familiar objects incorrectly (0% correct). Crucially, each child, regardless of her/his respective group membership, only witnessed the out-group as more reliable. For example, in the 100% vs. 0% condition if a child was a member of the Red Group, they would view four instances of the Blue Group (out-group) speaker correctly labeling the four familiar objects, whereas the Red Group (in-group) speaker would produce four inaccurate labels. After each trial, the experimenter paused the video and said, "The girl/boy in the red/blue shirt called it a _____, and the girl/boy in the red/blue shirt called it a _____. What do you think it is called?" The experimenter pointed to the screen to make it clear which actor provided which label for the familiar object. Once the child provided their own label, they were asked to point to whom they thought labeled the object correctly. Half the participants were Blue Group members, and therefore viewed the Red Group speakers being more reliable. Within this group (Blue Group participants) half saw the reliable Red Group speaker on the right side of the screen and half saw the reliable speaker on the left. The same counterbalancing measures were repeated for Red Group participants.

Film 2 (100% v. 25% condition) was identical in structure to Film 1 except one actor (out-group member) named all four familiar objects correctly (100% correct), and the other actor (in-group member) named one out of four familiar objects correctly (25% correct). On the trial in which both actors overlapped and provided accurate labels, each actor provided the same accurate label. Following each

video clip and at the end of the four accuracy trials, the experimenter asked the same questions as in Film 1. The particular trial in which both actors correctly named the object varied randomly across participants.

In Film 3 (75% v. 0% condition), one actor (out-group member) named three out of four familiar objects correctly (75% correct), and the other actor (in-group member) named all four familiar objects incorrectly (0% correct). On the trial in which both actors provided inaccurate labels, they provided different inaccurate labels. Following the video clips the experimenter asked the same questions as in Film 1 and 2. For half the participants the actor on the right was 75% correct, and for half the participants the actor on the left was 75% correct. The trial in which both actors provided inaccurate labels varied randomly across participants

Film 4 (75% v. 25% condition), one actor (out-group member) named three out of four familiar objects correctly (75% correct), and one actor (in-group member) named one out of four objects correctly (25% correct). In this condition, the actors' correct or incorrect labels never overlapped (i.e., there were no clips in which both actors made errors or in which both actors were correct). The same questions were asked as in Films 1, 2, and 3. The position of the clip in which the mostly accurate (75% correct) actor made an error and the mostly inaccurate (25% correct) actor provided the correct label varied randomly across participants.

Finally, at the end of the task, children were asked two questions to assess their thinking about group membership. First, "In the whole game, which group do you think was better at naming the objects, the Blue Group or the Red Group?" The order of these forced choice alternatives was counterbalanced across participants.

And second, “Which group do you belong to?” This was asked to ensure that children maintained an understanding of their own group membership throughout the task.

Results

As in Experiment 1, I first report children’s performance on the two separate group membership assessments. In addition, children’s responses to the “Which group was better” question are discussed. I then address children’s performance during accuracy trials. Next, children’s responses to the three test questions are analyzed and compared to chance performance. Finally, I examine the two series of Explicit Judgment questions separately to address any differences between the two.

Performance on group membership assessments. Children performed very well on both the initial and final group membership assessments. At the beginning of the task, all children accepted their group membership (Blue or Red), and accurately identified which group they belonged to. In addition, all children were able to correctly identify and label novel individuals as being members of “their” group or the “other” group. At the end of the task, each child was able to accurately identify which group they belonged to. Finally, children’s responses were not significantly different from chance, $t(19) = 1.902$, $p > .05$, when asked which group was better at naming the objects throughout the entire game.

Performance during accuracy trials. Across all four conditions, each participant was able to easily recognize the familiar objects chosen, and subsequently correctly identify which speaker was providing the accurate label for each object. Therefore, no children were excluded from analysis.

Comparisons to chance for the three test questions. Results for children's performance on the three test questions and comparisons to chance performance are presented in Table 3. Proportion correct scores and comparisons to chance for the Explicit Judgment (EJ), Ask, and Endorse questions in all four conditions were calculated. EJ scores were obtained from the final series of EJ questions, asked at the end of each condition. Children's performance did not differ from chance levels on all three question types across all four conditions.

Performance on both sets of explicit judgment questions. The results of children's responses to the first series of EJ questions, asked immediately following the 4 familiarization trials, are presented in Table 4. Overall, children performed well, accurately identifying the more reliable speaker above chance levels in 3 out 4 conditions: 100% vs. 0% correct, 100% vs. 25% correct, and 75% vs. 0% correct. However, children's performance was not above chance levels in the 75% vs. 25% correct condition.

The results of children's performance on the second set of EJ questions, asked at the end of each condition, are displayed in Table 5. Across all four conditions, children's performance did not significantly differ from chance.

Discussion

The results of Experiment 2 support and extend the findings of Experiment 1 that the existence of group membership alters children's ability to selectively trust more accurate individuals. Similar to Experiment 1, children again demonstrated a propensity for categorizing themselves and others into the novel Blue and Red

Groups, as evidenced by their perfect performance on group membership assessments. However, the results of Experiment 2 suggest that when children are asked to compare a reliable out-group member to an unreliable in-group member, group membership dominates and preschoolers' ability to use prior reliability as a cue for selective trust is completely disoriented.

Children's performance on Explicit Judgment (EJ) measures demonstrates a discernible impact of group membership. Immediately following the presentation of reliability information, children performed relatively well on the explicit measures, struggling to identify the more reliable speaker only in the most ambiguous condition (75% vs. 25% correct). Although this is the most difficult ratio to track, children's lack of success is striking if one considers that these EJ probes were asked no more than a minute after children witnessed the relative reliability of the two speakers. Furthermore, without group membership (Pasquini et al., 2007) and when comparing in-group members (Exp. 1) children have no problem tracking this information. Thus, the impact of group membership is already evident in children's diminished ability to immediately process ambiguous prior reliability information.

Although children's initial reliability tracking was only slightly weakened, their lack of success during the second set of EJ questions demonstrates an even more powerful effect of group membership. Children were incapable of consistently identifying the more accurate versus the less inaccurate informant when asked at the end of each condition, merely 5-8 minutes removed from the initial reliability information. Thus, by the end of the experiment, any reliable versus unreliable speaker distinction had apparently disappeared. This is somewhat surprising, as

previous research has demonstrated that 4-year-olds have little difficulty maintaining an unreliable versus reliable categorization over that short period of time (Birch et al., 2007; Koenig & Harris, 2004; Koenig et al., 2004; Pasquini et al., 2007). Further, several studies have shown that children will maintain these distinctions for up to 24 hours to a week later (Corriveau & Harris, 2009a). In contrast, when attempting to evaluate dissimilar group members, 4-year-olds appeared incapable of sustaining such a representation. A similar lack of success on EJ probes was reported by Corriveau and Harris (2009b) with regard to children's evaluations of familiar versus unfamiliar teachers' testimony. These authors offer the possible interpretation that children are prone to misremember information that conflicts with their initial schemas. In other words, if children enter a given learning situation with the expectation that in-group members are more reliable, which is evidenced by prior research (Kinzler et al., in press), then they might be more likely to later misremember the speaker's actual history of accuracy/inaccuracy.

The main goal of Experiment 2 was to investigate children's evaluations of a reliable out-group member and an unreliable in-group member. 4-year-olds' chance performance on Explicit Judgment (post-test), Ask, and Endorse questions in all four conditions suggests that their selective trust systems were completely disoriented when confronted with this decision. Surprisingly, in a powerful display of in-group favoritism, children were unable to override a speaker's group status in order to trust a more reliable out-group member. This pattern even occurred when the out-group speaker was entirely accurate and was compared to an entirely inaccurate in-group speaker (i.e., 100% versus 0% correct condition).

In sum, these results suggest that when presented with the scenario of evaluating a reliable out-group member against an unreliable in-group member, children's ability to selectively trust more reliable speakers is severely disoriented. Furthermore, children's ability to track the prior reliability of two speakers is initially only slightly diminished, but over a relatively short period of time is considerably disrupted. Thus, while 4-year-olds are completely capable of tracking and using previous accuracy to drive current decisions of whom to trust, these results suggest that they are incapable of overriding in-group status in order to trust a more reliable out-group member over a less reliable in-group member.

--------- FIGURES 1 AND 2 ABOUT HERE ---------

General Discussion

The overarching goal of both Experiments 1 and 2 was to assess the influence of group membership on young children's demonstrated ability to selectively trust and learn from more reliable speakers. Would a speaker's in- vs. out-group status take precedence over their prior history of reliability when children are choosing whom to trust? Taken together, the results from Experiments 1 and 2 support three main conclusions. First, the addition of minimal social categories, regardless of in- or out-group comparisons, desensitizes children's reliability tracking systems. Second, when comparing the relative trustworthiness of two in-group members, children require unambiguous prior accuracy information in order to make an accurate evaluation. And, third, children were unable to override initial in-group preferences in order to trust a more reliable out-group informant. I will discuss each finding in turn.

Across both experiments, children demonstrated a diminished capability to track and identify a more reliable speaker. Recall that when comparing the relative reliability of in-group members (Exp. 1), children, while initially successful, performed at chance on the second set of Explicit Judgment (EJ) questions (5-8 minutes after witnessing evidence of reliability) in 2 out of the 4 conditions. In addition, children proved incapable when asked to track and compare the relative reliability of an in-group and an out-group speaker (Exp. 2). These findings stand in stark contrast to previous selective trust research that does not include group membership, which demonstrates children's relative sensitivity and skill when tracking this sort of information (Birch et al., 2007; Koenig & Harris, 2004; Koenig et al., 2004; Pasquini et al., 2007). Thus, the inclusion of minimal group categorization, at the very least, disrupted children's reliability tracking systems, which could significantly alter children's patterns of selective trust.

The impact of group membership, however, is not limited to disorienting children's reliability tracking systems. In addition, these results suggest that a speaker's group status alters children's ability to use prior reliability as a cue for trustworthiness. In Experiment 1, children were asked to compare the relative reliability of two in-group members across 4 accuracy/inaccuracy ratios, which varied in ambiguity. In the least ambiguous scenarios, children had little difficulty using prior reliability as a cue for selective trust; however, in the more ambiguous conditions, in which the unreliable actor provided at least one instance of accuracy, children did not choose to learn from the more reliable speaker at a rate greater than chance. Thus, in order for preschoolers to categorize in-group members as more or

less trustworthy, the prior reliability evidence must be unequivocal; otherwise, children appear unwilling to dismiss a relatively inaccurate in-group speaker as untrustworthy.

A potential explanation for children's diminished ability to selectively trust one in-group member over another (when presented with ambiguous prior reliability information) comes from research conducted by Bigler et al. (1997) investigating children's intergroup attitudes. When Bigler and colleagues placed 6- to 9-year-old children into minimal groups (blue vs. yellow shirts), they found that children evaluated in- and out-group members in a pattern consistent with an "in-group homogeneity" effect – children were more likely to attribute positive traits to all members of their in-group. Perhaps, when presented with at least one example of an in-group member demonstrating numerous instances of prior reliability, which all participants saw in Experiment 1, children are likely to assume that other in-group members will exhibit a similar level of reliability (i.e., in-group homogeneity). If this were the case, it would be less crucial for children to track prior instances of accuracy and inaccuracy in order to make a reliable vs. unreliable distinction between the two in-group members. This explanation fits well with the finding that ambiguous prior reliability information was not sufficient to label one of the in-group speakers as more reliable, and consequently not sufficient to selectively trust one speaker over the other. In sum, if a speaker shares in-group status, it seems that children weigh prior reliability information differently, requiring clear evidence of unreliability to dismiss an in-group member as untrustworthy.

It is crucial to note that there is a second plausible interpretation for children's diminished success in Experiment 1 (in-group to in-group comparison). It could be that children demonstrated less sensitive reliability tracking because of an increased difficulty with source monitoring based on the fact that both speakers wore matching blue t-shirts. If children had difficulty tracking and recalling who was the more reliable speaker due to the informants' similar appearance, their selective trust would have been desensitized irrespective of the inclusion of group membership. Past research has typically used different shirt colors, often utilizing them as referents in selective trust questions (e.g., Who do you think was better at naming the objects, the girl in the red shirt or the girl in the blue shirt?) (Corriveau & Harris, 2009a; Fusaro & Harris, 2008; Pasquini et al., 2007). However, both children's success on EJ probes in the less ambiguous conditions in Experiment 1 and lack of success on EJ probes in Experiment 2 provide convergent support that the results of the in-group to in-group comparison can not be fully explained by an increased difficulty with source monitoring. Namely, children were able to track and differentiate the prior reliability information of two speakers wearing the same shirt color when that information was unambiguous (e.g., the 100% v. 0% correct condition), and children displayed a similar pattern of poor performance on EJ probes in Experiment 2 even though shirt color varied. Thus, the inclusion of group membership is a stronger explanation for children's reduced ability to track a speaker's prior history of accuracy and inaccuracy.

A third major finding from this study is that preschoolers were unable to selectively trust a more reliable out-group speaker over a less reliable in-group

speaker. Such a strong effect of group membership was not expected, but children's lack of selective trust extended to all three measures (Explicit Judgment, Ask, and Endorse). Across all four conditions, children failed to explicitly identify, direct questions towards, and endorse the more reliable out-group speaker. Furthermore, children were unable to correctly identify which group had been more accurate throughout the entire task, demonstrating a marked influence of including an in- to out-group comparison. Why might children's selective trust have been disoriented in this scenario? There are several possible explanations.

First, it could be that children approach in- to out-group comparisons with preconceived notions about the trustworthiness of group members. Recent research has demonstrated that children consider certain social categories to be more trustworthy in the absence of prior reliability information. Specifically, children will trust a familiar over an unfamiliar speaker (Corriveau & Harris, 2009b), an adult over a child (Jaswal & Neely, 2006), and a native-accented speaker over a foreign-accented speaker (Kinzler, Corriveau, & Harris, in press). Therefore, it is reasonable to think that children initially consider in-group members as more trustworthy, and that this belief might alter how children processed incoming prior reliability information during this task. Creating a scenario where, in order to be deemed trustworthy, out-group members had to provide overwhelming evidence of prior reliability. Whereas, in-group members were only required to provide minimal evidence of reliability before they were trusted. However, children were unable to trust an out-group speaker even when they had demonstrated a perfect track record of accuracy when labeling familiar objects. Thus, it is sensible to conclude that shared

in-group status proved too powerful a cue for children to moderate their initial preferences for in-group members and to selectively trust a more reliable out-group speaker.

A second plausible explanation for the lack of trust in out-group speakers is that children were not making a dispositional attribution about in-group members based on the presented information. Previous research has demonstrated that individuals are more likely to attribute negative behaviors performed by in-group members to situations rather than dispositions, whereas the opposite is true for out-group members (i.e., the fundamental attribution error) (Hamilton & Trolier, 1986). During the familiarization phase of Experiment 2, children witnessed in-group members performing rather poorly when labeling familiar objects. It may be the case that children are less likely to interpret these mistakes as representative of an enduring dispositional trait such as unreliability or untrustworthiness. Rather, children may have interpreted the in-group speakers' errors as them "pretending" or "joking", creating a scenario where in-group members could make mistakes and still be deemed trustworthy. A similar pattern of differential thinking about equivalent prior reliability information was found by Corriveau and Harris (2009b). Children explained a given speaker's mistakes when labeling well-known objects differently depending on their social category – familiar vs. unfamiliar. Familiar speakers' mistakes were more often explained as "pretending", while unfamiliar speakers' mistakes were attributed to "ignorance". Taken together, these results suggest that social category and group membership does influence children's processing of equivalent prior reliability information, which likely alters children's selective trust.

There is an alternative interpretation for the strong effect of minimal group membership found in this study, which warrants discussion. By design, this experiment began by placing children into one of two novel groups and subsequently asking a series of "group reinforcement" questions. These questions were intended to clarify and differentiate the similar in-group against the dissimilar out-group. However, it could be argued that children interpreted the experimenter's initial focus on minimal groups as a cue for the desired response strategy to be used during the experiment, prompting children to attend to group membership and answer the selective trust probes with this in mind. If this were the case, children might selectively trust in-group members because this is what they believe the task calls for, giving less attention to the prior reliability information. This interpretation is doubtful for two reasons. First, besides the initial set of group reinforcement probes, all other aspects of the task suggest that children should attend to prior accuracy information (e.g., explicit judgment questions). Second, in previous research investigating social categories and selective trust, children, although at first influenced by a speaker's social category, were capable of moderating these initial preferences in the presence of prior reliability information – trusting a more reliable informant over a less reliable but originally preferred informant (Corriveau & Harris, 2009b; Jaswal & Neely, 2006). In these studies, both prior reliability and social category information was present, and children chose to base their trustworthiness evaluations on a speaker's prior history of accuracy/inaccuracy. Thus, it is unlikely that children were simply disregarding the prior reliability information and choosing based on group membership. Nonetheless, future research should attempt to replicate these findings

without the use of group reinforcement probes, which would hopefully remove any experimenter influence on children's response patterns.

Another direction for future research is to address why children were unable to trust out-group speakers. Based on this study, it is clear that when compared to a less reliable in-group speaker, children default to trusting members of their own group. However, is this because out-group members are automatically categorized as untrustworthy, or because they are simply untrustworthy when compared to in-group members? Future research could investigate whether children would selectively trust a previously reliable out-group member in the absence of a direct in-group comparison, as it is unlikely that children are entirely unwilling to trust out-group speakers. Furthermore, future research should investigate the threshold at which an out-group member becomes reliable. Perhaps, children require more instances of out-group reliability than were provided in this study, and with a more pronounced difference between the unreliable in-group and the reliable out-group, children may prove capable of moderating their selective trust in favor of more reliable speakers, as previous research has demonstrated (Corriveau & Harris, 2009b; Jaswal & Neely, 2006). Finally, it is possible that these findings are age specific, and that as children grow older they become more sensitive to prior reliability as a viable cue for trustworthiness. A similar developmental shift was found by Corriveau and Harris (2009b): 3-year-olds were unable to overcome their initial preference for familiar speakers to trust an unfamiliar but more reliable speaker. Therefore, studies should be designed to explore whether older children (5-year-olds) still appear unwilling to trust more reliable out-group members.

In summary, these findings support three major conclusions. First, the inclusion of group membership alters preschoolers' reliability tracking mechanisms. Second, when children evaluate the relative trustworthiness of two in-group members, one speaker must demonstrate clear evidence of prior inaccuracy to be labeled as untrustworthy. And, third, preschoolers appear incapable of overriding their initial in-group preferences to selectively trust a more reliable out-group speaker. Taken together, these results suggest that children's decisions of whom to trust are inextricably linked with their social cognition. Children clearly consider a speaker's group membership when weighing instances of prior reliability, thus substantially altering their evaluations of others' trustworthiness. A strong interpretation of these results argues that group status (in- vs. out-) outweighs a prior history of reliability, making out-group members unlikely candidates for children to learn from. This finding may have broader practical implications for educational settings, in which any given teacher will frequently be an out-group member to at least one student. And, consequently, under some circumstances, children may prove more likely to question the accuracy of an out-group teacher. However, more research with the use of pre-existing groups in natural educational settings, in addition to a more diverse sample, is necessary before strong conclusions about the practical implications of these findings can be made. What can be concluded is that it is critical for adults to demonstrate care when communicating information about others' in- and out-group statuses. As this information, at the very least, might influence whether an individual is viewed as a viable source of knowledge, and, at the most, determine whether they can be trusted.

References

Aboud, F.E. (1988). *Children and Prejudice.* New York: Blackwell.

Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R. (2006). Nature and Nurture in Own-Race Face Processing. *Psychological Science,* 17, 159-163.

Baron, A., & Banaji, M. (2006). The Development of Implicit Attitudes: Evidence of Race Evaluations from Ages 6 and 10 and Adulthood. *Psychological Science, Vol. 17,* 53-58.

Brewer, M.B. (1979). In-Group Bias in the Minimal Intergroup Situation: A Cognitive-Motivational Analysis. *Psychological Bulletin,* 86, 307-324.

Bigler, R. S., Jones, L.C., & Lobliner, D.B. (1997). Social Categorization and the Formation of Intergroup Attitudes in Children. *Child Development, Vol. 68, No. 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, Vol. 72,* 1152-1162.

Birch, J.A., Vauthier, S.A., & Bloom, P. (2007). Three- and Four-year-olds spontaneously use others' past performance to guide their learning. *Cognition, 107,* 1018-1034.

Brewer, M.B., & Silver, M. (1978). Ingroup bias as a function of task characteristics. *European Journal of Social Psychology,* 8, 393-400.

Buchan, R.N., Johnson, J.E., & Croson, R. (2006). Let's get personal: An international examination of the influence of communication, culture, and

social distance on other regarding preferences. *Journal of Economic Behavior & Organization,* 60, 373-398.

Cameron, J.A., Alvarez, J.M., Ruble, D.N., & Fuligni, A.J. (2001). Children's lay theories about ingroups and outgroups. *Personality and Social Psychology Review,* 5, 586-591.

Chow, V., Poulin-Dubois, D., & Lewis, J. (2008). To see or not to see: Infants prefer to follow the gaze of a reliable looker. *Developmental Science, 11,* 761-770.

Clement, F., Koenig, M.A., & Harris, P.L. (2004). The Ontogenesis of Trust. *Mind & Language,* 19, 360-379.

Corriveau, K., & Harris, P.L. (2009a). Preschoolers continue to trust a more accurate informant 1 week after exposure to accuracy information. *Developmental Science,* 12, 188-193.

Corriveau, K., & Harris, P.L. (2009b). Choosing your informant: weighing familiarity and recent accuracy. *Developmental Science,* 12, 426-437.

Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences,* 7, 173-179.

Diesendruck, G., & haLevi, H. (2006). The Role of Language, Appearance, and Culture in Children's Social Category-Based Induction. *Child Development,* 77, 539-553.

Dion, K.L. (1973). Cohesiveness as a determinant of in-group—out-group bias. *Journal of Personality and Social Psychology,* 28, 163-171.

Guth, W., Vittoria, Levati, M., & Ploner, M. (2008). Social Identity and trust—An experimental investigation. *The Journal of Socio-Economics, 37,* 1293-1308.

Harris, P.L. (2007). Trust. *Developmental Science,* 10:1, 135-138.

Hamilton, D.L., & Trolier, T.K. (1986). Stereotypes and stereotyping: An overview of the cognitive approach. In J.F. Dovidio & S.L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 127-163). Orlando, FL: Academic Press.

Faulkner, P. (2002). On the Rationality of Our Response to Testimony. *Synthese, 131,* 353-370.

Fenson, L., Dale, D.S., Reznick, J.S., Bates, E., Thal, D.J., Pethick, S.J., Tomasello, M., Mervis, C.B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59,* pp. i-185.

Fusaro, M., & Harris, P.L. (2008). Children assess informant reliability using bystanders' non-verbal cues. *Developmental Science,* 11, 771-777.

Jaswal, V.K. (2004). Don't believe everything you hear: Preschoolers' sensitivity to speaker intent in category inductions. *Child Development,* 75, 1871-1885.

Jaswal, V.K., Neely, L.A. (2006). Adults don't always know best: Preschoolers' use past reliability over age when learning new words. *Psychological Science,* 17, 757-758.

Jaswal, V.K., & Malone, M.S. (2007). Turning Believers into Skeptics: 3-Year-Olds' Sensitivity to cues to Speaker Credibility. *Journal of Cognition and Development,* 8(3), 263-283.

Katz, P., & Kofkin, J. (1997). Race, gender, and young children. In S.S. Luthar, J.A. Burack, D. Chicchetti, & J. Weisz (Eds.). *Developmental Psychopathology:*

Perspectives on Adjustment, Risk, and Disorder (pp. 51-74). New York, NY, US: Cambridge University Press.

Kelley, D., Quinn, P., Slater, A.M., Lee, K., Gibson, A., Smith, M., Ge, Liezhong, & Pascalis, O. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science, 8,* F31-F36.

Kinzler, K.D., & Spelke, S.E. (2005). The effect of language on infants' preference for faces. Poster presented at the Biennial Meeting of the Society for Research in Child Development, Atlanta, GA.

Kinzler, K.D., Shutts, K., DeJesus, J., & Spelke, E.S. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition,* 27, 623-634.

Kinzler, K.D., Corriveau, K.H., & Harris, P.L. (in press). Children's selective trust in native-accented speakers. *Developmental Science.*

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

Koenig, M.A., & Echols C.H. (2002). Infants' understanding of false labeling events: the referential roles of words and the speakers who use them. *Cognition,* 87**,** 179-208.

Koenig, M.A., Clement, F., & Harris, P.L. (2004). Trust in Testimony. Children's Use of True and False Statements. *Psychological Science,* 15, 694-698.

Koenig, M.A., & Harris, P.L. (2004). The Role of Social Cognition in Early Trust. *Trends in Cognitive Sciences.* Vol. 9, No. 10, 457-459.

Koenig, M.A., & Harris, P.L. (2005). Preschoolers Mistrust Ignorant and Inaccurate Speakers. *Child Development, Vol. 76, No. 6,* 1261-1277.

Koenig, M.A. & Harris, P.L. (2007). The Basis of Epistemic Trust: Reliable Testimony or Reliable Sources? *Episteme,* 264-284.

Kowalski, K., & Lo, Y. (2001). The Influence of Perceptual Features, Ethinic Labels, and Sociocultural Information on the Development of Ethnic/Racial Bias in Young Children. *Journal of Cross-Cultural Psychology,* 32, 444-455.

Maccoby, E.E., & Jacklin, C.N. (1987). Gender segregation in childhood. In E.H. Reese (Ed.), *Advances in child development and behavior (*Vol. 20, pp. 195-216). New York: Academic Press.

Meissner, C.A., & Brigham, J.C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law,* 7, 3-35.

Nurmsoo, E., & Robinson, J.E. (2009). Identifying unreliable informants: do children excuse past inaccuracy? *Developmental Science,* 12:1, 41-47.

Park, B., & Rothbart, M. (1982). Perception of out-group homogeneity and levels of social categorization: Memory for subordinate attributes of in-group and out-group members. *Journal of Personality and Social Psychology,* 42, 1051-1068.

Pasquini, E.S., Corriveau, K.H., Koenig, M., & Harris, P.L. (2007). Preschoolers Monitor the Relative Accuracy of Informants. *Developmental Psychology, Vol. 43, No. 5,* 1216-1226.

Patterson, M. M., & Bigler, R.S. (2006). Preschool Children's Attention to Environmental Messages About Groups: Social Categorization and the Origins of Intergroup Bias. *Child Development, Vol. 77,* 847-860.

Pea, R.D. (1982). Origins of verbal logic: spontaneous denials by two- and three-year-olds. *Journal of Child Language,* 9, 597-626.

Poulin-Dubois, E., & Chow, V. (2009). The Effect of a Looker's Past Reliability on Infants' Reasoning About Beliefs. *Developmental Psychology, Vol. 45, No.6,* 1576-1582.

Quinn, P.C., Yahr, J., Kuhn, A., Slater, A.M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception,* 31, 1109-1121.

Rothbart, M., Davis-Stitt, C., & Hill, J. (1997). Effects of arbitrarily placed category boundaries on similarity judgments. *Journal of Experimental Social Psychology,* 33, 122-145.

Scofield, J. & Behrend, D.A. (2008). Learning Words from reliable and unreliable speakers. *Cognitive Development, 23,* 278-290.

Spelke, E.S. & Kinzler, K.D. (2007). Core Knowledge. *Developmental Science, 10:1,* 88-96.

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

Table 1

Stimuli Used in Accuracy Trials for Experiments 1 and 2.

Condition	*Familiar Object*	*Accurate Label*	*Inaccurate Label*
100% versus 0%	*Bear*	*"That's a bear"*	*"That's a shoe"*
	Bowl	*"That's a bowl"*	*"That's a dog"*
	Shoe	*"That's a shoe"*	*"That's a flower"*
	Book	*"That's a book"*	*"That's a ball"*
100% versus 25%	*Phone*	*"That's a phone"*	*"That's an apple"*
	Ball	*"That's a ball"*	*"That's a cookie"*
	Doll	*"That's a doll"*	*"That's a lion"*
	Bottle	*"That's a bottle"*	*"That's a clock"*
75% versus 0%	*Brush*	*"That's a brush"*	*"That's a plate"*
			"That's a cup"
	Cup	*"That's a cup"*	*"That's a duck"*
			"That's a towel"
	Hat	*"That's a hat"*	*"That's a penny"*
			"That's a truck"
	Boat	*"That's a boat"*	*"That's a tree"*
			"That's a nose"
75% versus 25%	*Fish*	*"That's a fish"*	*"That's a balloon"*
	Clock	*"That's a clock"*	*"That's a car"*
	Spoon	*"That's a spoon"*	*"That's a key"*
	Shovel	*"That's a shovel"*	*"That's a fork"*

Table 2

Stimuli Used in Test Trials for Experiments 1 and 2.

Condition	*Novel Object*	*Actor 1 label*	*Actor 2 label*
100% versus 0%	*Yellow column*	*"That's a dax"*	*"That's a rib-rub"*
	Orange Diablo toy	*"That's a liff"*	*"That's a wug"*
	Paint roller	*"That's a kern"*	*"That's a danu"*
	Rubber clamps	*"That's a neem"*	*"That's a gabber"*
100% versus 25%	*Thread wheel*	*"That's a fep"*	*"That's a gobi"*
	Purple koosh toy	*"That's a toma"*	*"That's a lorg"*
	3-hole punch	*"That's a merval"*	*"That's a zav"*
	Large paper cutter	*"That's a cham"*	*"That's a roke"*
75% versus 0%	*Foam tower*	*"That's a plick"*	*"That's a mogo"*
	Can opener	*"That's a feppin"*	*"That's a nevi"*
	Plastic arm sling	*"That's a niddy"*	*"That's a terval"*
	Plastic w/koosh toy	*"That's a riff"*	*"That's a norp"*
75% versus 25%	*Vid game controller*	*"That's a gog"*	*"That's a cloe"*
	Ball drop toy	*"That's a dint"*	*"That's a sup"*
	Yellow screwdriver	*"That's a fage"*	*"That's a tive"*
	Two compasses	*"That's a bove"*	*"That's a lut"*

Table 3

Proportion Correct and Comparisons With Chance for Experimental Measures: Explicit Judgment (EJ), Ask, and Endorse questions (Exp. 1 and 2).

Question	Exp. 1 (n = 22)		Exp. 2 (n= 20)	
	Proportion	*t*(21)	Proportion	*t*(19)
	100% vs. 0% correct condition			
EJ (Post-Test)	.86 (.22)	4.26***	.58 (.33)	0.75
Ask	.72 (.26)	3.36**	.56 (.28)	0.89
Endorse	.74 (.24)	3.95***	.51 (.27)	0.37
	100% vs. 25% correct condition			
EJ (Post-Test)	.79 (.26)	2.61*	.60 (.34)	0.89
Ask	.58 (.30)	1.05	.55 (.32)	0.62
Endorse	.64 (.28)	1.97	.54 (.30)	0.50
	75% vs. 25% correct condition			
EJ (Post-Test)	.65 (.31)	2.01	.62 (.33)	1.07
Ask	.51 (.30)	0.15	.58 (.33)	0.90
Endorse	.56 (.25)	0.93	.58 (.27)	1.10
	75% vs. 0% correct condition			
EJ (Post-Test)	.70 (.29)	1.88	.60 (.34)	0.89
Ask	.64 (.24)	2.24*	.59 (.29)	1.03
Endorse	.73 (.25)	3.58**	.59 (.33)	1.07

Standard errors are in parentheses
* p < .05. ** p < .01. *** p < .001.

Table 4

Number (Maximum = 1) of "Good" Responses to First and Second Explicit Judgment Questions and of "Better" Responses to Third Explicit Judgment Question (Post-Familiarization).

Question	Exp. 1 (n = 22)		Exp. 2 (n= 20)	
	Mean	SD	Mean	SD
100% vs. 0% correct condition				
Mean number of good judgments for 100% correct	0.95***	0.21	0.95***	0.22
Mean number of good judgments for 0% correct	0.05***	0.21	0.05***	0.22
Mean number of better judgments for 100% correct	0.95***	0.21	0.95***	0.22
100% vs. 25% correct condition				
Mean number of good judgments for 100% correct	0.95***	0.21	0.80**	0.41
Mean number of good judgments for 25% correct	0.09***	0.29	0.25	0.44
Mean number of better judgments for 100% correct	0.91***	0.29	0.80***	0.41
75% vs. 25% correct condition				
Mean number of good judgments for 75% correct	0.91***	0.29	0.80**	0.41
Mean number of good judgments for 25% correct	0.14***	0.30	0.20**	0.41
Mean number of better judgments for 75% correct	0.91***	0.29	0.75	0.44
75% vs. 0% correct condition				
Mean number of good judgments for 75% correct	0.82***	0.39	0.75	0.44
Mean number of good judgments for 0% correct	0.09***	0.29	0.20**	0.41
Mean number of better judgments for 75% correct	0.82***	0.39	0.80**	0.41

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 5

Number (Maximum = 1) of "Good" Responses to First and Second Explicit Judgment Questions and of "Better" Responses to Third Explicit Judgment Question (Post-Test).

Question	Exp. 1 (n = 22)		Exp. 2 (n= 20)	
	Mean	SD	Mean	SD
100% vs. 0% correct condition				
Mean number of good judgments for 100% correct	0.82**	0.40	0.55	0.51
Mean number of good judgments for 0% correct	0.18**	0.40	0.40	0.50
Mean number of better judgments for 100% correct	0.86***	0.35	0.60	0.50
100% vs. 25% correct condition				
Mean number of good judgments for 100% correct	0.73*	0.46	0.60	0.50
Mean number of good judgments for 25% correct	0.27*	0.46	0.40	0.50
Mean number of better judgments for 100% correct	0.77**	0.43	0.60	0.50
75% vs. 25% correct condition				
Mean number of good judgments for 75% correct	0.73*	0.46	0.60	0.50
Mean number of good judgments for 25% correct	0.32	0.48	0.35	0.49
Mean number of better judgments for 75% correct	0.68	0.48	0.60	0.50
75% vs. 0% correct condition				
Mean number of good judgments for 75% correct	0.68	0.48	0.60	0.50
Mean number of good judgments for 0% correct	0.32	0.48	0.35	0.49
Mean number of better judgments for 75% correct	0.68	0.48	0.60	0.50

* $p < .05$. ** $p < .01$. *** $p < .001$.

Figure Captions

Figure 1. Overall proportion of times children chose the more accurate informant on *Ask* trials. The solid black bar represents data from Pasquini et al. (2007) (no group membership). The solid red bar represents data from Experiment 1 (in-group to in-group comparison). Finally, the solid blue bar represents data from Experiment 2 (in-group to out-group comparison).

Note: * $p < .05$

Figure 2. Overall proportion of times children chose the more accurate informant on *Endorse* trials. The solid black bar represents data from Pasquini et al. (2007) (no group membership). The solid red bar represents data from Experiment 1 (in-group to in-group comparison). Finally, the solid blue bar represents data from Experiment 2 (in-group to out-group comparison).

Note: * $p < .05$

Figure 1 **Ask Trials**

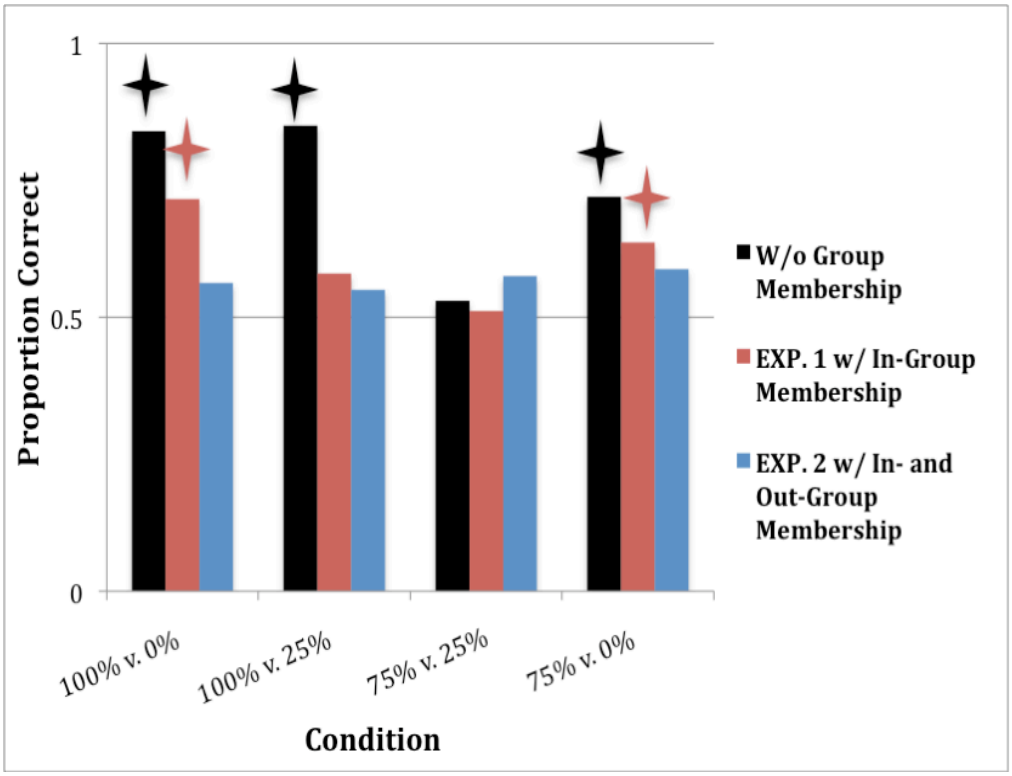

Figure 2 **Endorse Trials**