

Hostile Attribution of Intent and Aggressive Behavior: A Meta-Analysis

Bram Orobio de Castro, Jan W. Veerman, Willem Koops, Joop D. Bosch, and Heidi J. Monshouwer

A meta-analytic review was conducted to explain divergent findings on the relation between children's aggressive behavior and hostile attribution of intent to peers. Forty-one studies with 6,017 participants were included in the analysis. Ten studies concerned representative samples from the general population, 24 studies compared nonaggressive to extremely aggressive nonreferred samples, and 7 studies compared nonreferred samples with children referred for aggressive behavior problems. A robust significant association between hostile attribution of intent and aggressive behavior was found. Effect sizes differed considerably between studies. Larger effects were associated with more severe aggressive behavior, rejection by peers as one of the selection criteria, inclusion of 8- to-12-year-old participants, and absence of control for intelligence. Video and picture presentation of stimuli were associated with smaller effect sizes than was audio presentation. Staging of actual social interactions was associated with the largest effects. The importance of understanding moderators of effect size for theory development is stressed.

INTRODUCTION

Most recent studies of social cognition and aggressive behavior are based on Dodge's influential model of social information processing (Dodge, 1986) and a reformulation of this model by Crick and Dodge (1994). The models propose that to react appropriately to social situations, social information has to be processed in an orderly fashion: (1) the information has to be encoded accurately; (2) the encoded information has to be represented correctly; (3) an interaction goal needs to be specified; (4) response alternatives have to be generated; (5) these response alternatives have to be evaluated and, from these responses, an optimal response has to be selected; and (6) the selected response has to be enacted. According to Dodge (1986), atypical processing during any of these steps may lead to aggressive behavior. A fair number of studies have shown that aggressive behavior is associated with deviations in encoding, goals, response generation, response evaluation, and enactment. Results of empirical studies on representation are, however, less consistent. The goal of the present meta-analytic review was to clarify these inconsistent findings concerning representation.

Representation of social information has been studied almost exclusively in the sense of representation of other people's intentions. In a typical study, participants are presented with a number of social situations in which another person hinders them and are asked why this person hindered them. If children attribute hostile intent to the other, they are more likely to respond unkindly or aggressively than if they represent the other's intentions as benign or accidental (e.g., Dodge, Pettit, McClaskey, & Brown, 1986). According to Dodge (1980), one of the reasons

why aggressive children act aggressively is that they attribute hostile intentions to others more often than do other children. This phenomenon has become widely known as "hostile attribution of intent" or "hostile attribution bias." In this article, the first term is used, because the word "bias" has a normative connotation that is not necessarily appropriate (cf. Garbarino, 1998; Trachtenberg & Viken, 1994).

Hostile attribution of intent has become an important element in theoretical accounts of the development of aggressive and antisocial behavior. It has been proposed that well-known risk factors for the development of aggressive behavior problems, such as rejection by peers (Coie & Dodge, 1988) and harsh parenting (Weiss, Dodge, Bates, & Pettit, 1992), predispose children to attribute hostile intent, particularly if their cognitive capacities to process these experiences are limited (Dishion, French, & Patterson, 1995; Dodge, 1993, 1999; Dodge, Lochman, Harnish, Bates, & Pettit, 1997). Hostile attributions of intent, in turn, are believed to cause aggressive behavior, instigate more problematic social interactions, and thereby limit nonaggressive interactions that could provide opportunities to learn prosocial behaviors (Dodge, 1999; Weiss et al., 1992). Thus, it is suggested that hostile attribution may be a key element in the development and persistence of behavior problems over time. Hostile attribution of intent is also interesting in a more general sense; that is, as an example of a "working model" of the world that is believed to be shaped by experience and to influence current social functioning (Dishion et al., 1995; Dodge, 1993). Last

but not least, hostile attribution of intent by aggressive children is considered a target for cognitive behavioral interventions to reduce behavior problems (Hudley & Graham, 1993; Lochman & Lenhart, 1993).

Notwithstanding generally favorable reviews (Akhtar & Bradley, 1991; Dodge, 1993), results of empirical studies on hostile attribution and aggression vary considerably. Some studies have reported small effects (e.g., Dodge, 1980; Dodge et al., 1986), others have reported large effects (e.g., Graham, Hudley, & Williams, 1992; Waldman, 1996), and still others have found no significant effects (Dodge et al., 1997; Matthys, Cuperus, & van Engeland, 1999). It is not clear why findings from different studies diverge to this extent. Given these divergent findings, however, it is hard to represent the evidence for the hostile attribution hypothesis adequately. The present study used meta-analytic techniques to assess differences in effect sizes between studies on hostile attribution of intent and aggressive behavior in children, and to explore which characteristics of children and methods account for differences between effect sizes.

In theory, a number of child characteristics may be expected to influence the strength of relations between aggressive behavior and hostile attribution of intent. The most influential factor is probably the variation in severity of participants' aggressive behavior. If hostile behavior is related to hostile attribution of intent, one would expect larger differences in hostile attribution between extremely aggressive children and nonaggressive children than between somewhat more aggressive children and somewhat less aggressive children.

A second child characteristic that may influence effect sizes is the kind of behavior problems on which children are selected or compared. The relation between hostile attribution and aggressive behavior is expected to be particularly strong for reactive aggression; that is, for children who frequently behave aggressively out of anger in reaction to a presumed threat (Crick & Dodge, 1996; Dodge, 1991; Dodge et al., 1997). The relation is thought to be weaker or absent for children whose aggressive behavior is predominantly proactive; that is, planned, instrumental, and "cold blooded." Even though reactive and proactive aggression are generally found to be highly correlated—approximately .60 in nonreferred samples (see, e.g., Dodge et al., 1997; Hubbard, Parker, Ramsden, & Smithmyer, 1998) to .71 in psychiatrically impaired boys (Steffens, 1999)—studies with predominantly reactively aggressive samples may be expected to find larger effect sizes than studies with predominantly proactively aggressive samples.

A closely related issue is the relation between ag-

gressive behavior and peer sociometric status. Rejection by peers and aggression are used as combined selection criteria in a number of studies, thus excluding not rejected, aggressive children and including rejected, but not severely aggressive children (e.g., Dodge, 1985). Given the presumed role of peer rejection as a cause for hostile attributions (Dodge, 1999), hostile attribution of intent may be expected to be stronger in children selected on a combined criterion of aggression and rejection than in children selected on aggression only.

There is a well-documented negative relation between aggressive behavior and verbal intelligence (e.g., Huesmann, Eron, Lefkowitz, & Walder, 1984). A number of hostile attribution studies have controlled for group differences in verbal intelligence through matching or analysis of covariance. The relation between hostile attribution and intelligence is complex, however (cf. Gomez & Hazeldine, 1996). On the one hand, as described above, limited cognitive capacity is seen as one of the elements that possibly causes hostile attribution. From this perspective, to control for intelligence when studying hostile attribution is somewhat awkward, much like studying differences between a rainy and a sunny day while controlling for the clouds. On the other hand, hostile attribution studies may require participants to be fairly verbally intelligent, so that they are able to understand the stimuli and give a nuanced response. If these requirements disproportionately influence participants who are verbally less sophisticated, controlling for differences in intelligence would be an adequate approach. Regardless of the appropriateness of controlling for intelligence, both perspectives predict that, other things being equal, effect sizes will be smaller in studies that control for intelligence than in studies that do not.

Another influential characteristic may be children's age. A number of studies suggest that hostile attribution of intent is negatively related to age (Crick & Dodge, 1996). Possibly, children come to appreciate the possibility that someone may cause harm without the intention to do so, only at a certain age (Olthof, 1980). It has been suggested that this understanding develops more slowly in aggressive children, causing comparatively more hostile attribution of intent only in the age range where nonaggressive children have already developed an understanding of unintentional harm, whereas aggressive children have not (Crick & Dodge, 1996).

Another child characteristic that is important to look at is gender. Given established gender differences in the incidence and development of aggressive behavior, it also seems important to study gender dif-

ferences in the relation between hostile attribution of intent and aggressive behavior.

In addition to child characteristics, effect sizes may also be related to methodological characteristics of studies. Experimental manipulation of social stimuli presented to participants has shown that the hostile attribution effect in aggressive children is stronger when participants imagine that they are personally involved in the situation (Dodge & Frame, 1982) and when they feel threatened by a peer (Dodge & Somberg, 1987) or frustrated by an unjustly lost computer game (Orobio de Castro, 2000). Priming of hostile intentions (Graham & Hudley, 1994) and playing aggressive games on a computer (Kirsh, 1998) lead to a short-term increase in hostile attribution of intent by children. The tendency to attribute hostile intent is, to a certain extent, context specific (Dodge et al., 1986). Differences in stimulus presentation, stimulus contexts, and number of stimuli may therefore be expected to influence children's involvement, feelings, and expectations, and thereby influence effect sizes.

Differences in response format may also influence effect sizes (Lochman, Lampron, & Rabiner, 1989). Studies that combined open-answered questions with multiple-choice formats generally found larger effects for open answers than for closed-response formats (Milich & Dodge, 1984; Orobio de Castro, Veerman, Koops, & Bosch, 2001).

Children may respond differently in different settings—that is, tested individually versus tested in groups—because different demand characteristics operate in these settings. In an individual setting, children may feel more tempted to give answers that they feel the interviewer wants to hear and may hesitate to give socially undesirable answers. The larger anonymity of a group setting may provide children with an opportunity to give less socially desirable answers, but may also leave more room for distraction from the task at hand.

Effect sizes may also vary due to differences in hostile attribution scoring. Certain studies base their hostile attribution score on the proportion of hostile intent answers (e.g., Erdley & Asher, 1996), whereas other studies report scores based on hostile attributions minus benign attributions (Dodge & Somberg, 1987), or an average of ratings made on hostile-to-benign rating scales (Hudley & Graham, 1993). The first procedure is sensitive to possible extreme answering tendencies in aggressive children, whereas the other procedures are not.

Finally, effect sizes may be moderated by the reliability of hostile attribution scores. Low reliability results in lower effect sizes, because the large measurement error of unreliable measures contributes to within-group variance (e.g., Orwin, 1994).

In short, the present study examined the relation between hostile attribution of intent and aggressive behavior, and tested whether the above-mentioned child and method characteristics influenced the strength of this relation. Child characteristics that were expected to influence effect sizes were severity and kind of behavior problems, sociometric status, intelligence, age, and gender. Methodological characteristics that were thought to influence effect sizes included personal involvement, stimulus presentation, context, number of stimuli, response format, setting, scoring, and reliability.

METHOD

Study Selection

The aim of this study was to include all empirical studies of relations between aggressive behavior and hostile attribution of a peer's intentions in children, conducted between January 1974 and January 1999. Aggressive behavior was defined as all behavior through which a child causes harm or hindrance to someone else. This broad definition included syndrome-level classifications (such as psychiatric classifications of disruptive behavior disorders), symptom-level classifications (such as starting fights), and dimensional constructs (such as aggressiveness or externalizing problems). Although a number of studies in this field used rejection by peers to assess behavior problems, we chose not to include studies in which rejection was the only indicator of behavior problems, because peer rejection does not necessarily result from a child's aggressive behavior, but may also result from antisocial behavior by a child's peers or a "mismatch" between children. Hostile attribution of intent was defined as the hostile attribution of intentions to a peer in social situations in which the peer's intentions are ambiguous or vary systematically over presented situations (e.g., partly ambiguous, partly benign, and partly hostile).

Study selection was conducted in two steps. First, a large set of studies was retrieved by searches in *PsychInfo* (<http://www.psycinfo.com>) and *Dissertation Abstracts Online* (www.umi.com). The searches were based on the following search criterion: ["antisocial behavio(u)r," "aggressive behavio(u)r," "aggressiveness," "violence," "behavio(u)r problems," "behavio(u)r disorders," "Conduct Disorder," or "Oppositional Defiant Disorder"] AND ["attribution," "hostility," "social cognition," "social perception," "social information processing," or "social problem solving"], and was limited to human participants in childhood or adolescence, and English, French, Dutch, and

German languages. The searches resulted in 394 titles. An additional 80 titles were retrieved from diverse sources such as references to retrieved studies in reviews (Akhtar & Bradley, 1991; Crick & Dodge, 1994; Dodge, 1985, 1993; Orobio de Castro, 2000; Yoon, Hughes, Gaur, & Thompson, 1999) and *Web of Science* (<http://www.isinet.com>), consultation with authors in the field, and conference proceedings and advertisements from the 1997 Society for Research in Child Development and the 1998 International Society for the Study of Behavioral Development biennial meetings. Thus, 474 studies were retrieved in total.

Second, a selection was made from the retrieved titles on the basis of abstracts (plus articles and additional information from authors in case of doubt). To be selected, studies were required to contain empirical measures of both aggressive behavior and attribution of intent in children. Additional requirements included the following:

1. Presented stimuli were standardized and concerned social interactions with peers. Studies presenting different information to different participants were not included, because—by definition—stimulus information could not be compared with other studies. This criterion prohibited the inclusion of interesting studies in which stimulus information was partly determined by participants (e.g., Dodge & Newman, 1981; Lochman, 1987; Lochman & Dodge, 1998). Studies that exclusively presented interactions with adults were also excluded. If possible, when both interactions with peers and interactions with adults were presented, effect sizes were based on peer interaction data only.
2. Part of the stimuli were ambiguous. Theoretically, group differences in hostile attribution of intent are to be expected, particularly for ambiguous stimuli. Whenever possible, therefore, effect size calculations were based on responses to ambiguous vignettes only. A number of studies reported scores based on a mixture of ambiguous and nonambiguous stimuli. In these cases, effect sizes based on this mixture were calculated. Effect sizes based on mixed vignettes did not differ from effect sizes based on ambiguous stimuli.
3. Hostile attribution of intent was operationalized as a reaction to specific stimulus situations. Studies of hostility as a general personality trait were not included.
4. Measurement of hostile attribution was not preceded by experimental manipulation. When

studies involved experimental manipulations—for example, manipulation of affect (Dodge & Somberg, 1987), priming (Graham & Hudley, 1994), or treatment (Hudley & Graham, 1993)—only premanipulation data were included in the meta-analysis.

Several child characteristics prohibited inclusion of studies into the meta-analysis:

1. Studies with Attention-Deficit Hyperactivity Disorder (ADHD) as the only selection criterion were not included, unless the ADHD group had high aggressive behavior scores as well.
2. Studies that used social competence as the only selection criterion were excluded. Low social competence was not considered an indicator of aggressive behavior, because social competence and aggression are not necessarily highly negatively correlated. An exception was made when items in the social competence measure clearly referred to aggressive behavior (e.g., “physical fighting”).
3. When status groups were distinguished on reactive and proactive aggression, only reactively and pervasively (both reactively and proactively) aggressive groups were included, because no relation was expected between purely proactive aggression and hostile attribution of intent.
4. Studies that compared clinically aggressive children to other clinical groups, but not to nonaggressive controls, were not included because the difference in aggressiveness between clinical groups could not be assessed reliably.
5. As explained above, studies that used rejection as the only independent variable were excluded from the meta-analysis. Studies that used selection criteria of rejection by peers in combination with measures of aggressive behavior were included, however.
6. One study was excluded because of irresolvable inconsistencies in reported statistics.

Based on these criteria, 41 studies with 6,017 participants in total were included in the meta-analysis. (A list of these studies and their main characteristics with details of effect size calculation is available from the corresponding author on request, as is a list of excluded studies.)

Coding of Study Characteristics

To investigate the hypothesized relations between study characteristics and effect size, six rele-

vant child characteristics and seven relevant method characteristics were assessed for each study. Child characteristics studied were severity of children's behavior problems; kind of children's behavior problems; and selection on children's sociometric status, controlling for children's intelligence, age, and gender. Method characteristics studied were stimulus presentation, stimulus contexts, number of stimuli presented, response format, setting, scoring procedure, and reliability.

Child characteristics. The six child characteristics were operationalized as follows:

1. The range of aggressive behavior in each study was coded for severity. Three types of studies were distinguished that spanned an increasing range in severity of aggressive behavior. The least extreme set of studies was labeled "nonreferred general." Studies in this set contained a single sample, from the general nonreferred population, and studied the relation between aggressive behavior and hostile attribution either as a correlation in the entire sample or by comparing the hostile attribution scores of the more aggressive and the less aggressive half of the sample. The second set of studies was labeled "nonreferred extremes." Studies in this set compared two samples from the nonreferred general population that differed extremely in aggressive behavior. In most studies in this set, a group of children with scores more than 1 *SD* below the sample mean was compared with a group with scores more than 1 *SD* above the sample mean. The third set of studies was labeled "nonreferred versus referred." This set contained all studies in which a nonreferred sample was compared with a sample referred to any type of intervention (special education, psychiatric care, prison, and so forth) for aggressive, disruptive, or antisocial behavior problems.

2. The kind of behavior problems studied was classified as either generally aggressive behavior or excluding proactive aggression. Because studies of hostile attribution only recently started to assess reactive and proactive aggression, many studies did not provide specific data concerning these two kinds of aggression. To assess the impact of reactive versus proactive aggression on effect sizes, it was therefore necessary to resort to a comparison between studies with generally aggressive children on the one hand, and studies that excluded proactively aggressive children on the other.

3. With regard to sociometric status, a distinction was made between studies that used a combined selection criterion of aggression and rejection by peers, and studies that selected on aggressive behavior regardless of sociometric status.

4. Further distinctions were made between studies that controlled for effects of (verbal) intelligence through matching or analysis of covariance and studies that did not do so.

5. Attempts were made to code the age of participants. Coding of age, however, was complicated by the fact that a large proportion of studies contained participants from a large range of ages, typically ages 6 to 12. As a result, age was classified by distinguishing among studies with participants who were 4 to 6, 6 to 12, 8 to 12, and over 12 years of age.

6. Gender differences were studied by noting the percentage of boys in each sample.

Method characteristics. To capture relevant methodological differences between studies seven methodological characteristics of studies were coded.

1. Five modes of stimulus presentation were distinguished: video presentation, audio presentation (read aloud by an experimenter or played from tape), text presentation of stories that were read by participants, picture presentation, and staging of an actual social interaction between the participant and a peer.

2. Social contexts of stimuli were coded following Dodge, McClaskey, and Feldman's (1985) taxonomy of social problem situations (provocation, peer group entry, and so forth). Categories were added to this classification if additional contexts were studied.

3. The number of social situations presented was also coded.

4. Response format was coded as open-answer or multiple choice/rating scales. If both response formats were used in a single study, the response format was coded as open-ended, and calculation of effect size was based on open-ended responses.

5. The setting of each experiment was coded as individual or in groups.

6. Calculation of hostile attribution scores was coded as either the proportion of hostile intent answers, or hostile attributions minus benign attributions or an average of ratings made on hostile-to-benign rating scales.

7. Cronbach's α reliability was noted, if available.

In addition to these characteristics, attempts were made to directly code participant's involvement in the presented situations. However, involvement could not be coded reliably on any of the variables that were constructed for this purpose. Indicators of involvement, such as the participant's perspective (e.g., first person, spectator), instruction (e.g., "imagine this happens to you"), stimulus content (e.g., "a game" versus "the highest level of your favorite game"), or participant's distress (e.g., ratings, self-report) were reported too globally and too idiosyncratically to be comparable across studies. (To save

space, additional study characteristics coded for exploratory purposes—for example, informants, measures of aggressive behavior, and location—are not described here. Information on these characteristics is available from the corresponding author on request.)

Data Analyses

Data analyses were conducted using Mullen's (1989) algorithms. All study outcomes were transformed to Fisher *Z*s. Fisher *Z* is similar to a correlation coefficient (*r*), corrected for nonlinearity of the correlation coefficient at extreme values. Fisher *Z* calculations were based on reported test statistics, or, if necessary, *t* tests conducted on reported means and standard deviations. All meta-analytic calculations were performed on Fisher *Z*s weighted by study sample size. To facilitate interpretation, Fisher *Z* scores were transformed to correlations for presentation in this text. The magnitude of correlations was interpreted following Cohen's (1977, as cited in Cohen & Cohen, 1983) conventions that correlations of .10, .30, and .50 correspond respectively to small, moderate, and large effect sizes.

Three meta-analytic questions were studied. First tested was whether the mean effect size was significantly larger than 0; that is, the likelihood of finding the observed effect sizes if no relation between hostile attribution and aggressive behavior existed. The robustness of this test was assessed by calculating a so-called fail-safe number, which is the number of hypothetical additional studies averaging null results required to make the overall effect non-significant. Second, hypothesized relations between effect size and study characteristics were tested. Third, the distribution of effect sizes was studied by means of tests for heterogeneity. A set of effect sizes is considered to be heterogeneous if the variation in effect sizes is so large that it is unlikely, $p < .05$, that the effect sizes represent one single population effect size. Because a heterogeneous group of effect sizes does not represent a single population effect size, a single estimate of a mean effect size is not very meaningful for heterogeneous sets of effect sizes. Heterogeneous sets of studies were therefore divided into homogeneous clusters of studies with similar study characteristics, and the mean effect size for each homogeneous cluster of studies was calculated.

The applied statistical analyses assume that effect sizes are independent. This assumption is generally problematic in meta-analyses because studies are rarely completely independent, if only because multi-

ple studies are conducted by the same authors in the same institutes.¹ To maximize independence of effect sizes, only one effect size was calculated for each study, with two exceptions: Lochman and Dodge (1994) compared groups of nonaggressive boys with moderately aggressive and severely violent boys in two age groups. Because of the scarcity of studies with severely violent boys all findings in this article were included as four effect sizes: nonaggressive versus moderately aggressive boys and nonaggressive versus severely violent boys in two age groups. Thus, in each age group, one study in the nonreferred extremes group and one study in the nonreferred versus referred group shared the same nonaggressive participants. Furthermore, Crick and Dodge (1996) reported on two age groups separately. Results from each age group were included as separate effect sizes.

In analyses of relations between effect size and study characteristics, the unit of analysis was studies rather than participants. The relatively small number of included studies in combination with a strong interdependence between study characteristics resulted in (nearly) empty cells for certain combinations of study characteristics. To prevent excessive influence of single studies with unique combinations of characteristics, at least two studies were required in each cell for tests of effects of study characteristics.

RESULTS

Study Characteristics

Forty-one studies with 6,017 participants in total were included in the meta-analysis. Twenty-seven effect sizes were published in peer-reviewed journals or monographs, three effect sizes were derived from unpublished data made available on request by authors of peer-reviewed articles concerning related topics, nine studies were presented in dissertations, and two studies were presented as posters in conferences. Thus, 34% of the effect sizes were derived from unpublished data. If the unpublished effects had been consistently smaller than the published effects, this might have been indicative of publication bias. The opposite effect was found, however, with larger unpublished effects than published effects, $Z = 2.53$, $p = .006$.

Thirteen out of twenty authors contacted for addi-

¹A clear interdependence between studies in the present meta-analysis was that Dodge co-authored 19 of the 41 studies, which included more than half of all participants. This might raise concerns that significant findings were not independent replications of earlier studies. Participation of Dodge as co-author, however, was not associated with larger effect sizes.

tional information about their studies kindly provided (part of) the requested information. All studies were coded by the first author and 13 randomly selected studies were double coded. Interrater agreement was over 90%, $\kappa > .80$, for all study characteristics, with the exception of participants' emotional involvement. The latter was therefore not included in further analyses, as explained in the method section. Frequency distributions of study characteristics are presented in Table 1.

Child characteristics. Concerning severity, 10 studies contained nonreferred general samples, 24 studies contained nonreferred extreme samples, and 7 studies contained nonreferred versus referred samples. In all but 4 studies, the kind of aggressive behavior children were compared on was general aggressive behavior. These 4 studies excluded predominantly proactively aggressive children. Sociometric status was a selection criterion in 14 studies; 13 of these concerned nonreferred extreme samples. Eight studies controlled for differences in (verbal) intelligence. With regard to gender, 20 studies included only boys, whereas only 1 study included only girls. The other studies all included both boys (38% to 84% of participants) and girls. Two studies included 4- to 6-year-old children, 14 included 6- to 12-year-old children, 17 included 8- to 12-year-old children, and 8 studies included children over 12 years of age.

Method characteristics. Twenty studies presented stimuli by audio, 13 studies used video, 2 studies used only pictures, 2 studies required participants to read stories themselves, and 2 studies staged an actual social interaction for participants. Information on presentation mode of 2 studies could not be retrieved. Many studies presented stimuli set in multiple social contexts. Contexts presented were being provoked or hindered (38 studies); peer group entry (6 studies); coping with social expectations (1 study); failure (2 studies); and being bullied (1 study), falsely accused (1 study), or threatened (1 study). All stimuli concerned interactions with peers, except in one study in which half the stimuli concerned interactions with adults. The number of stimulus vignettes presented to participants differed markedly among studies ($range = 1-18, M = 6.1, SD = 3.9$). An open-answer response format was used in 27 studies, whereas 14 studies used only multiple-choice or rating-scale formats. Children were tested in an individual setting in 33 studies, and in groups in 7 studies. No information on setting could be obtained for 1 study. Calculation of hostile attribution scores was based on the proportion of hostile intent answers in 21 studies. In another 16 studies, calculation was based on hostile attributions minus benign attributions or on an average of

ratings made on hostile-to-benign rating scales. Scoring information was not provided in 4 articles. Reliability information was provided for only 18 studies. For these studies, Cronbach's α varied from .53 to .90, with a mean of .75.

Relations between study characteristics. Study characteristics were not distributed independently over studies. Certain combinations of characteristics occurred more frequently than others, and some combinations of characteristics never occurred. For example, severity was strongly related to several study characteristics (as shown in Table 1), such as kind of problem behavior, control for intelligence, and participants' gender, $ps < .05$. The interdependence between study characteristics prohibited analyses of higher order interaction effects, because these would have involved (nearly) empty cells for rare combinations of characteristics.

Overall Effect Size

Figure 1 shows the distribution of effect sizes. Thirty-seven out of forty-one effects were in the hypothesized direction. The overall relation between hostile attribution of intent and aggressive behavior was highly significant, $Z = 11.25, p < .001$, with a weighted mean effect size of $r = .17$. This mean effect size should be interpreted with care, because effects differed considerably between studies, ranging from $r = -.29$ to $r = .65$. Effect sizes were significantly heterogeneous between studies, $\chi^2(40) = 188.37, p < .001$. The fail-safe number of 3,411 indicates that over 3,400 additional studies averaging null results would be necessary to make the overall combined probability level nonsignificant.

Moderators of Effect Size

Effect sizes were significantly related to severity of behavior problems, $Z = 3.78, p < .001$. Weighted mean effect sizes were $r = .13$ for the 10 studies that involved nonreferred general samples, $r = .18$ for the 24 studies that involved nonreferred extreme samples, and $r = .23$ for the 7 studies that compared referred to nonreferred samples. Within each severity classification, effect sizes differed considerably: for nonreferred general samples, $\chi^2(9) = 22.33, p = .008$; for nonreferred extreme samples, $\chi^2(23) = 121.69, p < .001$; and for nonreferred versus referred samples, $\chi^2(6) = 33.75, p < .001$.

Because of the strong relation between severity and effect size, analyses of the relations between study characteristics and effect sizes were conducted in two steps. First, analyses were conducted over all

Table 1 Moderators of Effect Size by Severity Classification

Characteristic and Level	Severity											
	All			Nonreferred General			Nonreferred Extremes			Nonreferred versus Referred		
	K	N	r ^a	K	N	r ^a	K	N	r ^a	K	N	r ^a
Child characteristics												
Kind												
General aggression	37	4,826	.19	9	2,236	.13	21	1,983	.24	7	607	.23
Proactive excluded	4	1,191	.08	1	114	(.22)	3	1,077	.06	—		
Sociometric status												
Not applicable	27	4,845	.14	9	2,236	.13	11	2,002	.12	7	607	.23
Rejected	14	1,172	.28	1	114	(.22)	13	1,058	.29	—		
Intelligence												
Controlled	8	5,215	.10	—			2	248	-.07	6	554	.18
Not controlled	33	802	.18	10	2,350	.13	22	2,812	.20	1	53	(.62)
Age group (years)												
4–6	2	112	.31	2	112	.31	—			—		
6–12	14	2,152	.10	4	943	.06	5	785	.10	5	424	.20
8–12	17	3,040	.22	2	1,060	.20	14	1,899	.22	1	81	(.36)
≥12	8	713	.12	2	235	.03	5	376	.15	1	102	(.25)
Gender												
Only boys	20	1,727	.22	2	173	.25	12	1,000	.24	6	554	.18
Both genders	20	4,272	.14	8	2,177	.12	11	2,042	.14	1	53	(.62)
Only girls	1	18	(.65)	—			1	18	(.65)			
Method characteristics												
Presentation												
Read by participants	2	327	.24	—			2	327	.24	—		
Audio/read aloud	20	2,850	.24	4	1,441	.18	12	1,226	.29	4	283	.31
Video	13	1,458	.09	5	661	.08	5	473	.05	3	324	.15
Pictures only	2	735	.02	1	248	(-.02)	1	487	(.05)	—		
Staged	2	57	.55	—			2	57	.55	—		
Not reported	2						2					
Context												
Provocation	38	5,329	.17	9	2,202	.14	23	2,573	.20	6	554	.18
No provocation	3	688	.10	1	148	(.06)	1	487	(.05)	1	53	(.62)
Stimuli												
≤5	19	1,640	.24	4	582	.13	12	828	.32	3	230	.21
>5	22	4,377	.14	6	1,768	.13	12	2,232	.12	4	377	.24
Response format												
Open answer	27	3,421	.14	4	708	.03	17	2,247	.14	6	466	.30
Multiple choice or rating scale	14	2,596	.20	6	1,642	.18	7	813	.28	1	141	(-.04)
Setting												
Individual	33	4,032	.17	7	1,334	.11	19	2,091	.19	7	607	.23
Group	7	1,837	.17	2	868	.18	5	969	.15	—		
Not reported	1			1								
Scoring												
Hostile	21	3,330	.19	6	1,588	.16	12	1,557	.19	3	185	.42
Mean/rating	16	2,261	.15	3	614	.08	11	1,464	.16	2	183	.30
Unclear	4			1			1			2		
Total	41	6,017	.17	10	2,350	.13	24	3,060	.18	7	607	.23

Notes: ^a = *r*s are transformations of mean Fisher Zs weighted by sample size; K = number of studies; ns = not significant; — = not applicable or not tested. If K = 1, *r* is presented in parentheses.

effect sizes. Second, to avoid confounding severity with other variables, separate analyses were conducted within each severity group.

Moderators of effect size over all studies. Mean effect

sizes by moderators and severity classification are listed in Table 1. Several child characteristics were related to effect size. Contrary to the hypothesis for kind of behavior problems, the mean effect size for

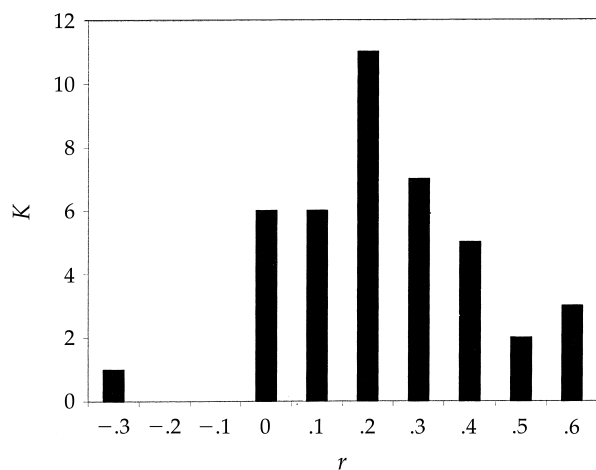


Figure 1 Distribution of effect sizes.

the four studies that excluded purely proactively aggressive children, $r = .08$, was smaller than for studies with generally aggressive participants, $r = .19$, $Z = 3.78$, $p < .001$. The mean effect size was larger for studies that included sociometric status as a selection criterion, $r = .28$, than for studies that selected solely on aggression, $r = .14$, $Z = 3.66$, $p < .001$. Studies that controlled for individual differences in intelligence had a smaller mean effect size, $r = .10$, than did studies that did not control for intelligence, $r = .18$, $Z = 2.54$, $p = .006$. Effect sizes for studies with participants who were 4 to 6, 6 to 12, and over 12 years of age were smaller than the effect size for studies with participants who were 8 to 12 years of age, with respective r s of .31, .10, .12, and .22, $Z = 2.79$, $p = .003$. With regard to gender, effects were larger for studies that included only boys, $r = .22$, than for studies with participants of both genders, $r = .14$, $Z = 1.89$, $p = .029$.

Effect sizes were also related to method characteristics. With regard to stimulus presentation, a nonsignificant mean effect size of $r = .02$ was found for studies that presented only pictures. Significantly larger effects were found for studies that presented stimuli on video, $r = .09$, $Z = 2.18$, $p = .015$, whereas larger effect sizes than for video were found in studies in which participants listened to stories, $r = .24$, or were required to read stories themselves, $r = .24$, $Z = 5.07$, $p < .001$. These studies, in turn, found smaller effects than the two studies that staged an actual interaction, $r = .55$, $Z = 2.25$, $p = .012$. The number of stimuli presented to participants was negatively related to effect size, $Z = 3.08$, $p < .001$. This finding may be explained by the strong relation between presentation mode and number of stimuli presented, with staged interaction consisting of only one situation, read stimuli consisting of multiple stimuli,

and video presentation consisting of the largest number of stimuli. Studies that scored the number of hostile intent responses to calculate the dependent variable found larger effects, $r = .19$, than studies that used a combination of hostile and benign responses, $r = .15$, $Z = 1.76$, $p = .039$. Effect sizes were not related to context, response format, setting, or (for the 16 studies that contained reliability information) reliability; respective probabilities, $Z = .60$, $p = .275$; $Z = .19$, $p = .423$; $Z = 1.50$, $p = .067$; and $Z = 1.30$, $p = .097$.

Moderators of effect size in nonreferred general samples. Most of the 10 studies in this group were designed with equal kinds of behavior problems and sociometric status, and none of the studies controlled for intelligence, precluding analyses of the effects of variation on these characteristics. With regard to age, studies with 4- to 6-year-old participants had larger effects, $r = .31$, than studies with 6- to 12-year-old participants, $r = .06$, $Z = 2.35$, $p = .009$. Studies that involved 8- to 12-year-old participants found larger effect sizes, $r = .20$, than studies with 6- to 12-year-olds, $r = .06$, $Z = 2.46$, $p = .007$, and studies with older children, $r = .03$, $Z = 2.29$, $p = .011$. Studies that involved both genders found smaller effects, $r = .12$, than studies that involved only boys, $r = .25$, $Z = 2.04$, $p = .021$.

Effect sizes were marginally larger for audio, $r = .18$, than for video presentation, $r = .08$, $Z = 1.53$, $p = .063$. Studies that presented only pictures had a nonsignificant mean effect size, $r = .02$. Studies that scored the hostile intent responses found larger effects, $r = .16$, than studies that averaged hostile and benign responses, $r = .08$, $Z = 2.29$, $p = .011$. Effect sizes in this set of studies were not related to context, number of stimuli presented, response format, or setting.

Moderators of effect size in nonreferred extreme samples. Several child characteristics in this group were related to effect size. The kind of behavior problems on which children were selected was influential. The three studies that excluded purely proactively aggressive children reported smaller effect sizes, $r = .06$, than did the other studies, $r = .24$, $Z = 5.33$, $p < .001$. Selection on sociometric status was associated with larger effect sizes, $r = .29$, than selection on aggression only, $r = .12$, $Z = 3.74$, $p < .001$. Two studies controlled for differences in intelligence, with a nonsignificant mean effect size, $r = -.07$, $Z = 5.29$, $p < .001$. Effect sizes were larger for children who were 8 to 12 years of age and older, $r = .22$, than for 6- to 12-year-olds, $r = .10$, $Z = 2.25$, $p = .012$. In this set, no study involved participants younger than 6. Gender was not related to effect size, $Z = 1.25$, $p = .11$.

Concerning stimulus presentation, the two studies that actually staged a real-life social interaction re-

ported larger effect sizes, $r = .55$, than did studies in which vignettes were read to or by children, respective r s of .29 and .24, $Z = 2.17, p = .015$. These, in turn, had larger effect sizes than studies that presented situations on video, $r = .05, Z = 3.23, p < .001$. Effect sizes were smaller for studies that presented a larger number of stimuli, $r = .12, Z = 3.72, p < .001$, which may again be explained by the fact that studies that used video presented more situations than did other studies. Open-answer response formats were associated with smaller effect sizes, $r = .14$, than were multiple-choice or rating-scale formats, $r = .18, Z = 2.16, p = .016$. Setting and scoring procedure were not related to effect size, $Z = 1.04, p = .15$ and $Z = .63, p = .27$, respectively.

Moderators of effect size in nonreferred versus referred samples. Seven studies compared hostile attribution of intent between nonreferred and referred samples. Because in this set all studies but one had the same characteristics for kind, sociometric status, intelligence, gender, context, response format, and setting, analyses of these moderators were not possible. The largest effect size of $r = .62$ differed significantly from all others, $Z = 3.54, p < .001$. This difference may be explained by a number of unique characteristics of that particular study, because it was the only one that did not control for intelligence, included both gen-

ders, and presented threatening rather than provocation contexts.

For the remaining characteristics, the following effects were found. Once again, effect sizes were associated with presentation. Larger effects were found for audio and read stimuli, $r = .31$, than for video, $r = .15, Z = 2.26, p = .006$. Effect size depended on scoring procedure. Larger effects were found when scores were derived from the number of hostile answers, $r = .42$, than when scores were derived from the number of hostile answers minus the number of benign answers, $r = .30, Z = 1.64, p = .025$. No relation between effect size and age or number of stimuli presented was found for this set of studies, $Z = .242, p = .41$ and $Z = .759, p = .22$, respectively.

Summary of moderators. Significant moderators of effect size are summarized in Table 2. The right-hand column of this table contains a summary of the effect of each moderator across study sets. Strongest evidence for the moderating effect of a variable is obtained when the moderating effect is replicated in different, independent sets of studies. Stimulus presentation was the only variable that consistently moderated effect size across the three sets of studies. Effect sizes were largest in staged interactions, smaller in studies in which stimuli were read to or by children, even smaller in studies in which situations were shown on video, and

Table 2 Significant Moderators of Effect Size by Severity Classification

Moderator	Severity Classification				Summary
	All	Nonreferred General	Nonreferred Extremes	Nonreferred versus Referred	
Child characteristics					
Sociometric status	Rejected > not applicable	—	Rejected > not applicable	—	+
Kind	General aggression > proactive excluded	—	General aggression > proactive excluded	—	+
Intelligence	Controlled < not controlled	—	Controlled < not controlled	Controlled < not controlled	++
Age group	8–12 > 4–6 and 6–12 and 12+	8–12 > 4–6 and 6–12 and 12+	8–12 > 6–12 and 12+	—	++
Gender	Boys > both genders	Boys > both genders	<i>ns</i>	—	+
Method characteristics					
Presentation	Staged > audio > video > pictures	Audio ≥ video > pictures	Staged > audio > video	Audio > video	+++
Context	<i>ns</i>	—	Provocation > no provocation	—	+
Situations	Few > many	<i>ns</i>	Few > many	<i>ns</i>	+
Response format	<i>ns</i>	<i>ns</i>	Multiple choice > open answer	—	+
Setting	<i>ns</i>	<i>ns</i>	<i>ns</i>	—	0
Scoring	Hostile > mean	Hostile > mean	<i>ns</i>	Hostile > mean	+

Note: — = not applicable or not tested; *ns* = not significant; +++ = consistent effect in all three sets; ++ = consistent effect, not tested in all sets; + = effect in one set; 0 = no effect.

smallest or absent in studies in which only pictures were shown. Other moderator effects could not be tested in all sets of studies, but were significant when they could be tested. These moderator effects were sociometric status (larger effects when aggressive children were also rejected by their peers), intelligence (smaller effects when intelligence was controlled for), and age (larger effects for 4- to 6-year-olds and 8- to 12-year-olds than for 6- to 12-year olds and children over 12). The effects of the remaining moderators were less clear. Effect size varied in part of the study sets with participants' gender (larger effects for boys), context (larger effects when provocation was included), response format (larger effects for multiple choice), number of stimuli (larger effects with fewer stimuli), and scoring (larger effects for hostile than for mean scores).

Homogeneous Clusters of Effect Sizes

Once the significant moderators of effect size were identified, they were used to explain the considerable heterogeneity in effect sizes between studies. To this end, studies with similar characteristics were grouped in clusters with homogeneous effect sizes. Because severity and presentation form were consistently related to effect size, clusters were formed first on the basis of these two variables. Other moderators were used to group studies only when combinations of these two variables did not result in homogeneous clusters. For parsimony, attempts were made to group as many studies as possible into as few clusters as possible. This resulted in seven homogeneous clusters of studies, containing 38 of the 41 effects in the analysis. These clusters are listed below and are depicted in Figure 2.

1. The first cluster contained six studies with 909 participants and involved video or picture presentation to nonreferred general samples. The homogeneous distribution, $\chi^2(5) = 9.99, p = .08$, of effect sizes in this cluster had a nonsignificant mean effect of $r = .06, Z = .44, p = .33$.

2. The second cluster contained five studies with 837 participants and involved video or picture presentation to nonreferred extreme samples. The homogeneous distribution, $\chi^2(4) = 8.28, p = .08$, of effect sizes in this cluster had a significant but small mean effect of $r = .10, Z = 2.02, p = .02$.

3. The third cluster contained four studies with 1,441 participants and involved audio or read presentation to nonreferred general samples. The homogeneous distribution, $\chi^2(3) = 5.13, p = .16$, of effect sizes in this cluster had a small mean effect of $r = .18, Z = 6.74, p < .001$.

4. The fourth cluster contained nine studies with 1,052 participants and studies involved audio or read presentation to nonreferred extreme samples. The homogeneous distribution, $\chi^2(8) = 1.45, p = .99$, of effect sizes in this cluster had a small to moderate mean effect of $r = .22, Z = 6.65, p < .001$.

5. Similarly, the fifth homogeneous cluster, $\chi^2(4) = 4.54, p = .34$, contained five studies involving 401 participants, with the same presentation and severity characteristics as in Cluster 4, but with larger effect sizes of $r = .44, Z = 9.16, p < .001$. Possibly, the larger effect sizes in this cluster resulted from inclusion of more extreme samples than in the Cluster 4 studies. Three of the five studies in Cluster 5 were conducted with similar instruments in the same region by (partly) the same authors (Graham & Hudley, 1994; Graham et al., 1992; Hudley & Graham, 1993). The authors of these studies stressed the im-

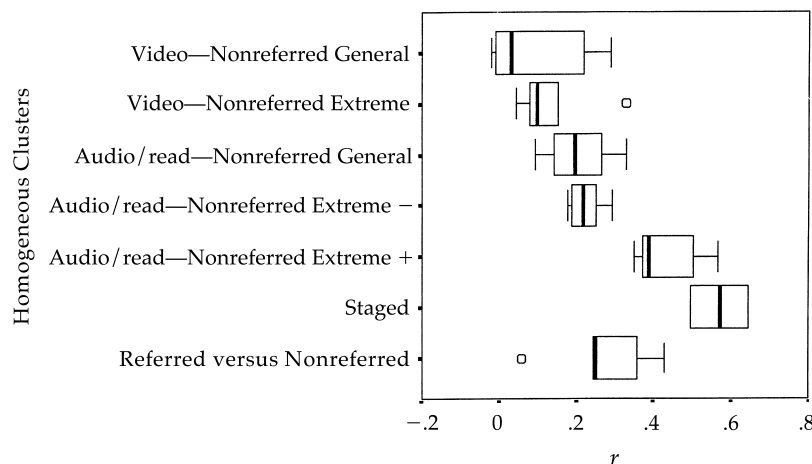


Figure 2 Homogeneous clusters of effect sizes. For each cluster the dark band shows the median, the bar indicates the quartile range, and the lines connect the extreme values. The two circles are considered outliers (see text).

pact of the deprived neighborhood in Los Angeles in which their predominantly low SES African American male participants lived on the formation of hostile attribution tendencies (Graham & Hudley, 1994). A fourth study (Hyatt, 1998) used *T* scores of 60 or higher as selection criteria on teacher ratings of aggressive behavior—a comparatively strict criterion (most studies that involve extreme nonreferred groups use a $M + 1 SD$ criterion). This explanation may also apply to the fifth study in this set (Vlerick, 1994), which compared extreme groups in the lowest and highest deciles of a locally devised aggression and social status questionnaire.

6. The sixth cluster contained the only two studies in which actual social interactions were staged for participants. This homogeneous cluster, $\chi^2(1) = .62$, $p = .43$, with 57 participants had the large mean effect size of $r = .55$, $Z = 4.12$, $p < .001$.

7. The seventh cluster consisted of five studies with 413 participants that compared referred to nonreferred samples and formed a homogeneous group, $\chi^2(4) = 7.01$, $p = .14$, with a moderate mean effect size of $r = .25$, $Z = 4.85$, $p < .001$.

Three effect sizes (7%) did not fit the above clusters. In addition, two effect sizes could not be clustered according to presentation mode, because presentation mode was not reported. The only significant negative effect in the meta-analytic database of $r = -.29$ obviously did not fit any cluster. This effect was one of four effects reported by Lochman and Dodge (1994), in the only study that derived scores from a factor analysis conducted on six combined samples of children who differed in age and behavior status. Possibly, extreme scores of children referred for severe behavior problems dominated this factor analysis, resulting in nonvalid scores for the less severely aggressive sample with the outlying effect size. This explanation is clearly highly speculative.

Both the study with the smallest and the study with the largest effect size involving referred children did not fit this cluster, probably because they differed from the other studies in this group on significant moderators of effect size. The only study in this group that found no effect (Matthys, Cuperus, & van Engeland, 1999) differed significantly from all other studies in that it used a combination of video presentation and control for intelligence, both of which were associated with smaller effect sizes. On the other side of the spectrum, the largest effect size in this group (Barrett, Rapee, Dadds, & Ryan, 1996) was found in the only study with referred participants that did not control for intelligence with threats as the stimulus context.

DISCUSSION

Studies of the relations between hostile attribution of intent to peers and aggressive behavior in children have obtained different results. Qualitative reviews have not adequately explained these divergent findings. In the present study, a meta-analysis was conducted to quantitatively integrate the available findings and to assess whether different findings could be explained by theoretically meaningful differences in study characteristics.

The use of meta-analysis to review the findings in this field entails potential weaknesses. A well-known threat to the validity of meta-analyses is publication bias. Nonsignificant study outcomes are generally underreported in the literature. If our attempts to retrieve all unpublished studies failed, disproportionately missed nonsignificant findings may have been obtained. However, even though it is not always possible to ensure that publication bias does not influence meta-analytic findings, there are good reasons to trust that publication bias did not distort the present findings. First, nearly all studies on social information processing concern multiple dependent variables, such as encoding and response generation, in addition to hostile attribution of intent. Because effects on these other dependent variables have consistently been significant, nonsignificant results on hostile attribution have frequently appeared in the literature as “fellow travelers” of these significant findings. Second, great lengths were taken to retrieve unpublished studies, resulting in a large proportion of effect sizes (34%) from unpublished data. Third, effect sizes from these unpublished studies were not smaller but larger than published effect sizes, suggesting that publication bias has not worked against publication of null findings in this field (but rather against studies with smaller samples). Fourth, even if, despite the above, publication bias did occur, it is unlikely to have influenced the robust main findings of the meta-analysis, because the fail-safe number indicated that the overall effect would only become nonsignificant if we failed to include over 3,400 additional studies that averaged null results. This seems highly unlikely.

A less obvious potential weakness of meta-analyses is the nonrandom distribution of study characteristics. Because study characteristics are not distributed independently, it may not be possible to unconfound the relations between different moderator variables and effect size. Some of our analyses of relations between study characteristics and effect size were indeed precluded by small cell sizes and confounding of study characteristics. For example, there were no studies on group settings in referred samples, and all staged in-

teractions involved nonreferred extreme samples. We chose to acknowledge these limitations imposed by the data set by analyzing only main effects within levels of severity, rather than to test for higher order interactions between moderator variables with nearly empty cells. As a result, not all effects of all moderator variables could be tested in each set of studies. Obviously, these limitations imposed by distributions of study characteristics are not unique to meta-analyses, but apply to any kind of review. Quantitative reviews have the comparative advantage of making these limitations explicit, as was done throughout this article.

The meta-analysis established a robust relation between hostile attribution of intent and aggressive behavior, and it showed large differences in effect sizes between studies that could mostly be explained by systematic analysis of study characteristics. Analyses of study characteristics revealed theoretically meaningful covariation of effect size with variations in both participant and task characteristics, with clear implications for theory and research.

Participant Status

Social information processing theory proposes a relation between hostile attribution of intent and aggressive behavior. This relation is expected to be stronger with larger differences between children in terms of severity of aggressive behavior problems, to be stronger for reactive and pervasive kinds of aggression than for proactive aggression, to be stronger when aggressive behavior is associated with rejection by peers, and to result in part from individual differences in (verbal) intelligence. Nearly all these relations were found in the meta-analysis.

Aggressive behavior was related to hostile attribution of intent. The strength of this relation differed considerably between studies, even though nearly all effect sizes were in the expected direction. As expected, effect sizes covaried with the range in severity of behavior problems studied. Obviously, the three-fold classification of severity used in this study was very crude. Because studies rarely reported scores on standardized measures of behavior problems, it was not possible to make more fine-grained comparisons between studies. Perhaps the additional variance in effect sizes can be explained by the more subtle differences in severity within each set of studies that were classified as "equally severe." Severity may have differed importantly between study locations; thus, differences in aggressive behavior between groups in "non-referred extreme" studies may be much more extreme in certain locations than in others—for example, in the cluster of large nonreferred extreme effect sizes.

Only wider use of standardized, internationally normed measures of behavior problems will enable more direct comparisons of participant status between studies. Such standardization is required for accumulation of knowledge in the field.

Effects were larger when children were selected on joint criteria of aggression and sociometric status. This finding is consistent with the proposition that rejection by peers contributes to the development of a disposition to make hostile attributions and vice versa. Nonetheless, sociometric status is a rather global measure of the interaction processes believed to operate in the formation of a tendency to attribute hostile intent to peers. It is, for example, not yet clear whether hostile attribution of intent is shown by rejected children with a small number of friendships, is shared by befriended dyads of rejected children, or is particularly strong in children who frequently change friends after arguments.

Controlling for intelligence was consistently associated with smaller effect sizes. There are two likely explanations for this effect. Individual differences in intelligence may distort measurements of hostile attribution in aggressive children, because their frequently limited (verbal) intelligence may impair understanding of stimuli and nuanced responding. Controlling for intelligence may have corrected for this distortion. Alternatively, individual differences in intelligence should not be considered a confounding factor, but may be an independent variable that causes individual differences in hostile attribution and aggressive behavior throughout development. As such, this factor explains inadequate learning of social problem-solving skills in less intelligent children. From this perspective, controlling for intelligence is not appropriate.

Information concerning the moderating effect of different kinds of behavior problems was scarce. Few studies provided data on kind of aggression, which limited the analyses to a rather crude comparison within the nonreferred extreme group between four studies that excluded purely proactively aggressive children and the remaining studies that presumably included these children. This comparison resulted in the opposite of the expected effect. Given the crude categorization of studies and the absence of nonreferred versus referred studies in this comparison, it is suggested that in this comparison, additional data are required.

No prediction was made with regard to a moderating effect of gender, but an interesting finding by itself was the relative absence of studies that included girls. One study included only girls, and only one of the seven clinical studies included any girls at all. Al-

though no consistent gender differences were found, it is not possible to conclude that the relation between hostile attribution of intent and aggression is the same in girls as in boys, because girls were so under-represented in the studies analyzed.

Detailed analyses of age effects were complicated by the wide age range sampled in a number of studies. Effects were stronger in studies with 4- to 6-year-olds and 8- to 12-year-olds than in studies that included 6- to 12-year-olds and children over 12 years of age. This finding is inconsistent with the "developmental lag" hypothesis, which proposes that aggressive children understand that "acts may be accidental, or have other effects than intended" later than do nonaggressive children. This hypothesis would predict an increase followed by a decrease in effect size with age, not the present wave-like pattern. In the absence of other hypotheses with regard to relations between age and attribution of intent, it is unclear why these age effects occurred. Note that the age effects only concerned variations in the magnitude of significant mean effect sizes. Significant relations between aggressive behavior and hostile attribution were found in children from 4 to over 12 years of age, suggesting that the developmental lag hypothesis does not explain the hostile attribution effect (or that this lag starts early and remains for a long period in development).

Task Characteristics

Assessment of task characteristics may clarify under what circumstances the hostile attribution effect occurs and under what circumstances it does not. To this end, the present study assessed relations between effect size and task characteristics. We believe that these analyses should not be interpreted as a "quest" for the task associated with the largest effect size, but rather as a means to improve theoretical understanding of conditions associated with individual differences in attribution of intent.

The relation between hostile attribution of intent and aggressive behavior was expected to be stronger with higher levels of personal involvement of participants in presented social situations. To test this hypothesis, unsuccessful attempts were made to code direct indicators of personal involvement. Unfortunately, little information concerning stimulus content was given in most research reports. Few articles presented sample stories, or even descriptions of more detailed themes of each situation than general descriptions such as "provocation" or "peer group entry." This lack of information precluded detailed comparisons of the situations presented to children. Given the assumed importance of contexts and involvement in

social information processing, presenting more information concerning contents and impact of situations in research reports (such as sample stories, or participant ratings of impact or emotion) seems essential.

Despite this setback, clear relations were found between effect size and stimulus presentation form, a task characteristic that may well be considered an indirect indicator of participants' personal involvement in tasks. In each of the severity groups, effects were consistently smaller, or even absent, when stimuli were presented on video or pictures, than when they were read to or by children. Even larger effects were found in the two studies that actually staged (seemingly) real-life interactions. The most likely theoretical account for this effect seems to be that children's personal involvement in stimuli depends on the way in which the stimuli are presented. As shown by Dodge and Frame (1982), the hostile attribution effect only occurs if children feel that they themselves are involved in the social interactions presented. Although video presentation is presumably meant to maximize children's involvement, it may, in reality, have the opposite effect, because seeing other children interact on video may make it hard for participants to imagine that they are part of the interaction themselves. Seeing a child you don't know in a blue T-shirt on a TV screen and following the instruction to "imagine it's you" may be quite difficult for children (and adults), thus, minimizing personal involvement. In contrast, when stimuli are read, children are directly addressed as partakers in the situation. Personal involvement will of course be largest if a participant is really personally involved in a staged interaction.

It is also possible that the large amount of irrelevant information in pictures and videos (e.g., actor's appearance and background features) distracts participants from the actual negative event, thereby reducing their involvement. Although group differences in the occurrence of such distraction may be interesting phenomena by themselves, they belong to the domain of encoding and should not distort measurement of representation. Yet an alternative explanation for the larger effects for audio than video presentation may be that only audio presentation directly activates cognitive schemas related to attribution of intent, if these schemas have partly been constructed by means of language; for example, through verbal instructions or self-talk.

Relations between effect size and several task characteristics were exploratorily assessed. Nearly all studies included provocation as a context. The largest effect size was found in the only study that described the social context of presented situations as "social

threats." Notwithstanding emphasis on the context specificity of social information processing in theoretical accounts (Crick & Dodge, 1994), little is known about the actual qualities of contexts that determine differences in hostile attribution of intent. The mere descriptor "provocation" in many reports is not very informative in this regard. The present analyses suggest certain important parameters of context. The importance of the events to the participant seems influential. Given the demonstrated influences of personal involvement and affective state on attribution of intent, one might, for example, expect a very different reaction to a mild, revocable provocation ("someone switching away from your favorite TV channel," quoted from the study with the smallest effect size in the referred study set) than to an irrevocable physical threat ("three kids you don't know walking toward you," quoted from the study with the largest effect size). It may well be that differences in importance of presented stimuli explain additional variance in study outcomes. In addition to importance, the extent to which situations are open to diverse interpretations may determine effect size. The largest openness to interpretation is present in the socially threatening situations associated with the largest effect size. In these situations, no actual harm has happened to the participant, but the harm that may be done if the peer's intentions are hostile is severe. Such open situations clearly leave more room for interpretation than detailed footage of an exact actual provocation.

The number of situations presented was only related to effect size in nonreferred extreme samples, in which larger effects were found when fewer situations were presented. The number of situations presented was closely related to stimulus presentation mode, with the largest number of situations presented on video, a smaller number presented on audio or read to or by participants, and only one situation being staged. Because number and presentation of situations were so closely related, their effects could not be distinguished in the present dataset. Independent variation of these two variables in future studies may clarify the role of the number of situations presented. Detailed analyses of existing data from studies that presented large numbers of situations may also clarify this issue, by testing a relation between the effect size for each separate situation, and the moment it was presented. If effect sizes consistently decline for each consecutive situation, this would present strong evidence for an effect of number of situations, independent of presentation mode. Furthermore, in combination with reliability calculations, an optimal test length could then be found as a balance between number of situations and reliability.

Response format and setting were also confounded, and strongly dependent on severity. Groupwise multiple-choice testing frequently occurred in less severe samples, whereas the most severe participants were almost exclusively tested individually by means of open-ended questions. Researchers probably assumed that groupwise multiple choice is the most efficient method to obtain large datasets in less severe groups, whereas the most severe groups would only be able to complete the tests if they were interviewed individually. Ironically, these assumptions could not be tested here, because their implementation resulted in the absence of individual open-ended tests in less severe groups and groupwise multiple choice tests in severe groups. Again, experimental manipulation of these study characteristics is required to test the assumptions upon which the study designs were based.

Rationales for using specific scoring procedures were not reported in any study. This is unfortunate. The core construct of a theory should be well defined to make valid operationalization possible. The variations in scoring procedures indicate that it is unclear whether hostile attribution of intent should be understood as "attribution of hostile intent, irrespective of attributions of benign intent," or as "attribution of hostile intent minus attributions of benign intent." The distinction is less academic than it may seem, because the two conceptualizations may lead to very different conclusions. If, for example, aggressive children tend to respond more extremely to questions in general (either by giving more explicit open answers, or by preferring the extreme-response categories in multiple choice), they would give more hostile and more benign responses to attribution questions. For the first of the above conceptualizations, this would result in higher hostile attribution scores; for the second conceptualization it would not. This phenomenon may explain the consistently larger effect sizes in studies that counted hostile intent responses. Clearly, neither of the two conceptualizations is wrong, but rationales for choosing either one of them are needed.

Reliability statistics were reported for less than half the studies. Unreported reliabilities are not necessarily low, but their absence signals an underappreciation of the importance of reliable measures in this field. Valid measurement requires reliable measures. Unreliable measures seriously reduce effect sizes. Assuming that mean reported reliability of .75 is an adequate estimate of the mean reliability over all studies, then true effect sizes would be, on average, $1/ (.75)^{1/2} = 1.15$ times larger than reported effect sizes. Furthermore, reliable measures are essential for diagnostic tests in clinical practice. If treatment decisions and evaluations are to

be based on hostile attribution scores; (as suggested by Akhtar & Bradley, 1991), insufficient reliability of tests can have serious consequences.

Social Information Processing

Relations between hostile attribution of intent and aggressive behavior are generally not studied for their own sake, but as elements of studies regarding social information processing. The present findings supported the basic tenets of social information processing models: (1) individual differences in behavior were associated with individual differences in processing of social information, (2) the strength of this association depended on characteristics of presented information, and (3) effects of child and task characteristics were in theoretically expected directions.

Notwithstanding these findings, several issues with regard to the role that hostile attribution of intent plays in social information processing remain unclear. First, hostile attribution of intent is nearly the only studied construct of the "representation" phase in the social information processing model. It is unclear whether the present findings can be generalized to other, less frequently studied elements of representation, such as representation of other's emotions or beliefs. Second, hostile attribution of intent is presented as an indicator of representation, but is generally an indicator of both representation and "encoding," the step that theoretically precedes representation in the social information processing model. Only one of the included studies actually controlled for encoding. All other effects may, in principal, be completely due to a relation between aggressive behavior and encoding that is reflected in the representation phase. A strict test of individual differences in representation of intent would somehow need to standardize or control for encoding of information for all participants. Third, the current analysis did not assess causality. Whether hostile attribution causes aggressive behavior or vice versa, whether there is a bidirectional effect, or whether a third factor (e.g., encoding) causes the relation is outside the scope of this article. Several studies have used experimental manipulations or structural equation modeling to test causal relations. Given the large number of these studies and the dependence of effects on child and task characteristics established in the present research, a meta-analytic integration of these studies may be worthwhile and feasible.

Aggressive Behavior

A broad definition of aggressive behavior was used to select studies for the meta-analysis. Included

studies contained data from different informants on diverse measures of aggressive behavior, ranging from observations to rating scales to police reports. Research on aggression has clearly shown low concurrent validity of broadly defined operationalizations of aggression between measures and informants. Given the generally low agreement between informants, the strength of relations between children's answers to hypothetical questions and aggressive behavior as indicated by teachers or parents in some studies is impressive.

Nonetheless, it seems likely that a considerable proportion of variation in effect sizes was due to differences in the assessment of aggressive behavior. Attempts to systematically explain this variation by coding informants and measures for each study were impracticable, due to the large variation in (combinations of) measures. Research into the kinds of aggressive behavior to which hostile attribution of intent is specifically related is clearly needed to resolve this issue.

Conclusions

The current meta-analysis quantitatively integrated the available findings concerning the relation between hostile attribution of intent and aggressive behavior in children and assessed whether divergent findings could be explained by theoretically meaningful differences in study characteristics. The analysis established a robust relation between hostile attribution of intent and aggressive behavior, and showed that the large differences in effect sizes between studies could be explained by systematic analysis of study characteristics. The analyses of study characteristics revealed theoretically meaningful sensitivity of the effect to variations in both participant and task characteristics, with clear implications for theory and research.

Future research in this area should include attempts to unconfound the different kinds of aggressive behavior, peer group status, and intelligence; longitudinal experimental studies to assess causal relations and development of hostile attributions and aggressive behavior; experimental manipulation of task characteristics and personal involvement of participants; and studies of psychometric qualities of social information processing tasks.

ACKNOWLEDGMENTS

The authors would like to thank the researchers of the included studies for generously providing them with additional information concerning the studies. They gratefully acknowledge Marinus van IJzendoorn,

John E. Lochman, Rolf Loeber, and the reviewers for their constructive advice.

ADDRESSES AND AFFILIATIONS

Corresponding author: Bram Orobio de Castro, Department of Developmental Psychology, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands; e-mail: b.castro@fss.uu.nl. Jan W. Veerman is at the University of Nijmegen, Nijmegen, The Netherlands; Joop D. Bosch is at Free University, Amsterdam; and William Koops and Heidi J. Monshouwer are also at Utrecht University.

REFERENCES

References marked with an asterisk (*) indicate studies included in the meta-analysis.

- Akhtar, N., & Bradley, E. J. (1991). Social information processing deficits of aggressive children: Present findings and implications for social skills training. *Clinical Psychology Review, 11*, 621–644.
- *Barrett, P. M., Rapee, R. M., Dadds, M. M., & Ryan, S. M. (1996). Family enhancement of cognitive style in anxious and aggressive children. *Journal of Abnormal Child Psychology, 24*, 187–203.
- *Bickett, L. R., Milich, R., & Brown, R. T. (1996). Attributional styles of aggressive boys and their mothers. *Journal of Abnormal Child Psychology, 24*, 457–472.
- *Brendgren, M., Bowen, F., Rondeau, N., & Vitaro, F. (1999). Effects of friends' characteristics on children's social cognitions. *Social Development, 8*, 41–51.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Coie, J. D., & Dodge, K. A. (1988). Multiple sources of data on social behavior and social status in the school: A cross-age comparison. *Child Development, 59*, 815–829.
- *Crick, N. C. (1995). Relational aggression: The role of intent attributions, feelings of distress, and provocation type. *Development and Psychopathology, 7*, 313–322.
- Crick, N. C., & Dodge, K. A. (1994). A review and reformulation of social information processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*, 74–101.
- *Crick, N. C., & Dodge, K. A. (1996). Social information processing deficits in reactive and proactive aggression. *Child Development, 67*, 993–1002.
- *Cuperus, J. M. (1997). *Sociale probleemoplossing bij kinderen met gedragsstoornissen* [Social problem solving in children with behavior disorders]. Unpublished doctoral dissertation, Utrecht University, The Netherlands.
- Dishion, T. J., French, D. C., & Patterson, G. R. (1995). The development and ecology of antisocial behavior. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Vol. 2. Risk, disorder, and adaptation* (pp. 421–471). New York: Wiley.
- *Dodge, K. A. (1980). Social cognition and children's aggressive behavior. *Child Development, 51*, 162–170.
- Dodge, K. A. (1985). Attributional bias in aggressive children. In P. C. Kendall (Ed.), *Advances in cognitive-behavioral research and therapy* (pp. 73–110). Orlando, FL: Academic Press.
- Dodge, K. A. (1986). A social information processing model of social competence in children. In M. Perlmutter (Ed.), *The Minnesota Symposium on Child Psychology: Vol. 18. Cognitive perspectives on children's social and behavioral development* (pp. 77–125). Hillsdale, NJ: Erlbaum.
- Dodge, K. A. (1991). The structure and function of reactive and proactive aggression. In D. Pepler & K. H. Rubin (Eds.), *The development and treatment of childhood aggression* (pp. 201–218). Hillsdale, NJ: Erlbaum.
- Dodge, K. A. (1993). Social-cognitive mechanisms in the development of conduct disorder and depression. *Annual Review of Psychology, 44*, 559–584.
- Dodge, K. A. (1999). *A biopsychosocial model of the development of chronic conduct problems in adolescence*. Manuscript in preparation.
- *Dodge, K. A., & Coie, J. D. (1987). Social information processing factors in reactive and proactive aggression in children's peer groups. *Journal of Personality and Social Psychology, 53*, 1146–1158.
- *Dodge, K. A., & Frame, C. L. (1982). Social cognitive biases and deficits in aggressive boys. *Child Development, 53*, 620–635.
- *Dodge, K. A., Lochman, J. E., Harnish, J. D., Bates, J. E., & Pettit, G. S. (1997). Reactive and proactive aggression in school children and psychiatrically impaired chronically assaultive youth. *Journal of Abnormal Psychology, 106*, 37–51.
- *Dodge, K. A., McClaskey, C. L., & Feldman, E. (1985). Situational approach to the assessment of social competence in children. *Journal of Consulting and Clinical Psychology, 53*, 344–353.
- Dodge, K. A., & Newman, J. P. (1981). Biased decision-making processes in aggressive boys. *Journal of Abnormal Psychology, 90*, 375–379.
- Dodge, K. A., Pettit, G. S., Bates, J. E., & Valente, E. (1995). Social information processing patterns partially mediate the effect of early physical abuse on later conduct problems. *Journal of Abnormal Psychology, 104*, 632–643.
- *Dodge, K. A., Pettit, G. S., McClaskey, C. L., & Brown, M. M. (1986). Social competence in children. *Monographs of the Society for Research in Child Development, 51*(2, Serial No. 213).
- *Dodge, K. A., & Price, J. M. (1994). On the relation between social information processing and socially competent behavior in early school-aged children. *Child Development, 65*, 1385–1397.
- *Dodge, K. A., & Somberg, D. R. (1987). Hostile attributional biases among aggressive boys are exacerbated under conditions of threats to the self. *Child Development, 58*, 213–224.
- *Dodge, K. A., & Tomlin, A. M. (1987). Utilization of self-schemas as a mechanism of interpretational bias in aggressive children. *Social Cognition, 5*, 280–300.

- *Erdley, C. A., & Asher, S. R. (1996). Children's social goals and self-efficacy perceptions as influences on their responses to ambiguous provocation. *Child Development, 67*, 1329–1344.
- *Feshbach, L. E. (1989). Aggression-conduct problems, attention-deficits hyperactivity, play, and social cognition in four-year-old boys. Unpublished doctoral dissertation, University of Washington, Seattle.
- Garbarino, J. (1998, July). *Making sense of "senseless" youth violence*. Paper presented at the XVth biennial meeting of the International Society for the Study of Behavioral Development, Berne, Switzerland.
- *Gibbins, C., & Craig, W. (1997, April). *Mapping the path to aggression: Validation of a social-cognitive model of childhood aggression*. Poster presented at the biennial meeting of the Society for Research in Child Development, Washington, DC.
- Gomez, R., & Hazeldine, P. (1996). Social information processing in mild mentally retarded children. *Research in Developmental Disabilities, 17*, 217–227.
- *Graham, S., & Hudley, C. (1994). Attributions of aggressive and nonaggressive African-American male early adolescents: A study of construct accessibility. *Developmental Psychology, 30*, 365–373.
- *Graham, S., Hudley, C., & Williams, E. (1992). Attributional and emotional determinants of aggression among African-American and Latino young adolescents. *Developmental Psychology, 28*, 731–740.
- *Guerra, N. G., & Slaby, R. G. (1989). Evaluative factors in social problem solving by aggressive boys. *Journal of Abnormal Child Psychology, 17*, 277–289.
- *Hart, M. T. (1993). *Social-cognitive processing in aggressive/withdrawn, aggressive/nonwithdrawn, and nondeviant children*. Unpublished doctoral dissertation, Texas A&M University, College Station.
- Hubbard, J. A., Parker, E. H., Ramsden, S. R., & Smithmyer, C. M. (1998, July). *Anger regulation processes in children's peer interaction*. Paper presented at the XVth biennial meeting of the International Society for the Study of Behavioral Development, Berne, Switzerland.
- *Hudley, C., & Graham, S. (1993). An attributional intervention to reduce peer-directed aggression among African-American boys. *Child Development, 64*, 124–138.
- Huesmann, P. L., Eron, L. D., Lefkowitz, M. M., & Walder, L. O. (1984). The stability of aggression over time and generations. *Developmental Psychology, 20*, 1120–1134.
- *Hyatt, L. F. (1998). *Moderating influences on the attributional bias towards hostile orientation in aggressive boys*. Unpublished doctoral dissertation, Hofstra University, Hempstead, NY.
- *Katsurada, E. (1995). *Preschoolers' hostile attribution, aggressive behavior, and relationships with their mothers' attributional style, parenting behavior, and affect*. Unpublished doctoral dissertation, Oregon State University, Corvallis.
- Kirsh, S. J. (1998). Seeing the world through mortal combat-colored glasses: Violent video games and the development of a short-term hostile attribution bias. *Childhood, 5*, 177–184.
- Lochman, J. E. (1987). Self- and peer perceptions and attributional biases of aggressive and nonaggressive boys in dyadic interactions. *Journal of Consulting and Clinical Psychology, 55*, 404–410.
- *Lochman, J. E., & Dodge, K. A. (1994). Social-cognitive processes of severely violent, moderately aggressive, and nonaggressive boys. *Journal of Consulting and Clinical Psychology, 62*, 366–374.
- Lochman, J. E., & Dodge, K. A. (1998). Distorted perceptions in dyadic interactions of aggressive and nonaggressive boys: Effects of prior expectations, context, and boys' age. *Development and Psychopathology, 10*, 495–512.
- Lochman, J. E., Lampron, L. B., & Rabiner, D. L. (1989). Format differences and salience effects in the social problem-solving assessment of aggressive and nonaggressive boys. *Journal of Clinical Child Psychology, 18*, 230–236.
- Lochman, J. E., & Lenhart, L. A. (1993). Anger coping intervention for aggressive children: Conceptual models and outcome effects. *Clinical Psychology Review, 13*, 785–805.
- *Matthys, W., Cuperus, J. M., & van Engeland, H. (1999). Deficient social problem-solving in boys with ODD/CD, with ADHD, and with both disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 38*, 311–321.
- *Milich, R., & Dodge, K. A. (1984). Social information processing in child psychiatric populations. *Journal of Abnormal Child Psychology, 12*, 471–490.
- Mullen, B. (1989). *Advanced BASIC meta-analysis*. Hillsdale, NJ: Erlbaum.
- Olthof, T. (1980). *Blame, anger, and aggression in children: A social-cognitive approach*. Unpublished doctoral dissertation, Katholieke Universiteit Nijmegen, Nijmegen, The Netherlands.
- Orobio de Castro, B. (2000). *Social information processing and emotion in antisocial boys*. Amsterdam: PI Research.
- *Orobio de Castro, B., Veerman, J. W., Koops, W., & Bosch, J. D. (2001). *Social information processing, emotion attribution, and emotion regulation in boys with antisocial behavior problems*. Manuscript submitted for publication.
- Orwin, R. G. (1994). Evaluating coding decisions. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis*. New York: Sage.
- *Quiggle, N. L., Garber, J., Panak, W. F., & Dodge, K. A. (1992). Social information processing in aggressive and depressed children. *Child Development, 63*, 1305–1320.
- *Sancilio, M. F., Plumert, J. M., & Hartup, W. W. (1989). Friendship and aggressiveness as determinants of conflict outcomes in middle childhood. *Developmental Psychology, 25*, 812–819.
- Steffens, H. (1999). *Kindspecifieke diagnostiek en behandeling van agressief gedrag [Individual assessment and treatment of aggressive behavior]*. Unpublished master's thesis, University of Amsterdam.
- *Steinberg, M. S., & Dodge, K. A. (1983). Attributional bias in aggressive adolescent boys and girls. *Journal of Social and Clinical Psychology, 1*, 312–321.
- Trachtenberg, S., & Viken, R. J. (1994). Aggressive boys in the classroom: Biased attributions or shared perceptions? *Child Development, 65*, 829–835.
- *Vlerick, P. (1994). The development of socially incompe-

- tent behavior in provocative situations. *Psychologica Belgica*, 34, 33–55.
- Waas, G. A. (1988). Social attributional biases of peer-rejected and aggressive children. *Child Development*, 59, 969–975.
- *Waldman, I. D. (1996). Aggressive boys' hostile perceptual and response biases: The role of attention and impulsivity. *Child Development*, 67, 1015–1033.
- *Weiss, B., Dodge, K. A., Bates, J. E., & Pettit, G. S. (1992). Some consequences of early harsh discipline: Child aggression and a maladaptive information processing style. *Child Development*, 63, 1321–1335.
- *White, D. Y. (1984). *Attributional bias in aggressive girls*. Unpublished doctoral dissertation, American University, Washington, DC.
- *Williams, R. L. (1998). *Social information-processing in lower and higher aggressive black adolescents*. Unpublished doctoral dissertation, Hofstra University, Hempstead, NY.
- *Yoon, J. S. (1998). *Social cognitive differences between aggressive-rejected and aggressive-nonrejected children*. Unpublished doctoral dissertation, Texas A&M University, College Station.
- Yoon, J., Hughes, J., Gaur, A., & Thompson, B. (1999). *Social cognition in aggressive children: A meta-analytic review*. Manuscript in preparation.
- *Zelli, A., Dodge, K. A., Lochman, J. E., Laird, R. D., & Conduct Problems Prevention Research Group. (1999). The distinction between beliefs legitimizing aggression and deviant processing of social cues: Testing measurement validity and the hypothesis that biased processing mediates the effects of beliefs on aggression. *Journal of Personality and Social Psychology*, 77, 150–166.