Children's Comprehension of Critical and Complimentary Forms of Verbal Irony

Jeffrey T. Hancock and Philip J. Dunham

Department of Psychology
Dalhousie University
Halifax, Canada

Kelly Purdy

Department of Psychology
McGill University
Montreal, Canada

The existing research on children's comprehension of verbal irony has focused exclusively on children's understanding of ironic criticisms. Two experiments examined 5- and 6-year-old children's ability to detect the nonliteral nature and intended meaning of both ironic criticism and ironic praise as depicted in short, videotaped stories. Considered together, the results from these experiments permit several conclusions: First, the data confirm earlier research suggesting that children's detection of nonliteral utterances and their interpretation of the speaker's pragmatic intent are separable components of early irony comprehension. Second, children's ability to detect ironic statements is asymmetrical across critical and complimentary forms of irony. Finally, although children more readily detect ironic criticisms, explicit echoic cues play an important role in facilitating uniquely their detection of ironic compliments. We discuss these results in the context of social pragmatic theories of early communicative development (e.g., Bruner, 1983; Tomasello, 1992, 1995) and with reference to a re-

cent allusional–pretense model of irony comprehension proposed for mature speakers (Kumon-Nakamura, Glucksberg, & Brown, 1995).

Verbal irony is a commonly used form of nonliteral speech in which the speaker's intended meaning is communicated indirectly. Although verbal irony can take many forms (Kumon-Nakamura, Glucksberg, & Brown, 1995), perhaps the simplest and most common form occurs when the speaker's intended meaning is the opposite of the literal statement. The specific forms such counterfactual statements can take include ironic criticisms and ironic compliments. An ironic criticism is a positive statement meant to convey a negative meaning, and an ironic compliment is a negative statement meant to convey a positive meaning. To illustrate, a girl who is playing basketball exclaims to her father, "Hey Dad, watch me play basketball." The girl then shoots the ball and misses the basket. The father replies, "You sure are a good basketball player" (ironic criticism). Conversely, if the girl makes the same initial statement to her father, then shoots the ball and makes the basket, the father might reply, "You sure are a bad basketball player" (ironic compliment). Note that in both forms, the ironic statement is counterfactual.

Although, in principle, ironic compliments and ironic criticisms are equally viable pragmatic options for a mature speaker, traditional theories emphasize that the usual purpose of verbal irony is to express a negative evaluation or attitude (e.g., Sperber & Wilson, 1981). Although available theories differ as to why this asymmetry exists, the general implications are (a) that ironic criticisms are more likely to be encountered in our daily discourse than ironic praise and (b) that, with some exceptions, ironic criticism tends to be detected and interpreted by mature conversationalists more readily than ironic praise (cf. Clark & Gerrig, 1984; Kreuz & Glucksberg, 1989; Sperber & Wilson, 1986). As Kumon-Nakamura et al. (1995) suggested,

> People can almost always express irony by using a positive assertion, such as "This is a terrific performance" when in fact the performance in question is terrible. The reverse, using a negative statement such as "This is a terrible performance" when the performance is actually quite good, seems anomalous (p. 11).

In this article, we focus on these issues during an early period of communicative development when children are just beginning to detect the nonliteral nature of ironic speech and to understand the pragmatic intent of the speaker. First, we note that the available developmental data concerned with the early comprehension of verbal irony are reasonably consistent. Dews and Winner (1997) reviewed a substantial amount of research concerned with how and when children's ability to comprehend verbal irony emerges during early development, and across a wide range of procedures and materials, they concluded that children are beginning to comprehend simple counterfactual forms of verbal irony between 5 and 6 years of

age (e.g., Ackerman, 1981, 1983; Andrews, Rosenblatt, Malkus, Gardner, & Winner, 1986; Demorest, Meyer, Phelps, Gardner, & Winner, 1984; Happe, 1993; Sullivan, Winner, & Hopfield, 1995).

Although the data are more limited, it is interesting to note in this context that children's early experiences with verbal irony may also be biased in the direction of ironic criticism. Dews, Winner, Nicolaides, and Hunt (as cited in Dews & Winner, 1997) estimated the frequency of ironic criticism and ironic praise in two types of television shows designed for children. These television shows averaged 2.75 instances of irony per 30-min segment. The utterances were predominantly critical in intent with only 6% judged as ironic praise. Although these data do not capture all of the discourse settings in which a child might encounter irony, the results imply that children, like adults, encounter ironic criticisms more frequently than ironic praise.

Given this background, our primary question of interest was whether children, during this early period of communicative development, display a bias similar to adults in their comprehension of ironic criticism and ironic praise. To the best of our knowledge, the literature concerned with children's early comprehension of irony has not addressed this question. Instead, researchers have focused exclusively on ironic forms with negative intent (i.e., ironic criticism). As such, an answer to this question fills a fundamental gap in our knowledge about the early development of irony comprehension.

A second question of interest in this study arises from measurement issues. In his early landmark research, Ackerman (1981, 1983) established the now-popular procedure of reading children various short stories that end in a terminal evaluative comment. The stories were constructed so that the final comment could be interpreted as either literal or ironic (as noted previously, all ironic endings were in the form of criticisms). His research indicated that detecting the literal versus nonliteral property of the final comment (e.g., was the statement accurate) and inferring the intent of the speaker (e.g., was the speaker angry or happy) were separable components of irony comprehension in young children. Although this dissociation is disputed in the context of adult comprehension research (cf. Dews & Winner, 1997; Gibbs, 1994), Ackerman's developmental observations are compelling. We therefore designed a procedure that permitted us to measure separately whether the children detected the nonliteral nature of a final comment (detection question) and whether they correctly inferred the intent of the speaker (intent question). The question of interest arising in the context of this measurement issue is whether any asymmetry observed in children's comprehension of ironic criticism and ironic praise would be consistent across both components of comprehension (detection and inference).

To address these questions, we focused specifically on 5- and 6-year-old children. As noted earlier, the existing literature establishes this as the period during which children are just beginning to comprehend ironic statements. As such, we

expected any differences in their sensitivity to ironic praise and ironic criticism to be most evident during this period when comprehension is well below ceiling levels of performance. Similarly, if, as has been suggested, children's detection of irony and their inferences about the pragmatic intent of the speaker are separable processes, this dissociation should also be most evident during this early transition period when performance is below ceiling levels.

EXPERIMENT 1

Method

Participants.   Participants were recruited from a database of volunteer families maintained in the Dalhousie University Infant Development Laboratory. Twenty-four English-speaking, middle-class, 5- to 6-year-olds participated in the study (16 girls, 8 boys). The 12 older children (7 girls, 5 boys) ranged in age from 74 to 78 months ($M = 76.17$, $SD = 1.75$). The 12 younger children (9 girls, 3 boys) ranged in age from 59 to 69 months ($M = 66.50$, $SD = 2.94$). The mean age for these groups combined was 71.33 months ($SD = 5.47$).

Materials.   Nine short stories were videotaped using four adult actors (two men, two women). Each story included two actors, A and B, engaged in an event familiar to a child, such as reciting the alphabet or playing baseball. In the initial scene, A boasted of a particular ability (e.g., I am a good basketball player). In the next scene, A either failed or succeeded at the task identified in the boast. If A succeeded at the task, B then offered either a literal or an ironic compliment; if A failed at the task, B offered either a literal or an ironic criticism. Nine such stories were filmed four times to include each of the four possible endings. Table 1 provides a sample of these stories as they were modified to create each of the four treatment conditions.

When filming the stories, the ironic and literal endings were identified for the actors. Although the existing literature suggests that children at this age are not particularly sensitive to the various prosodic cues that mark irony when the context is salient (cf. Milosky & Ford, 1997; Winner & Leekam, 1991), the actors were encouraged to employ natural changes in intonation that they would normally associate with the production of the literal and ironic statements.

Design.   Each participant watched 16 stories: Four ended with a literal criticism, 4 with an ironic criticism, 4 with a literal compliment, and 4 with an ironic compliment. Half of the children saw stories ending in criticisms first (i.e., literal or ironic), and the other half saw stories ending in compliments first (i.e., literal or ironic). The literal and ironic endings were presented in random order. As such, the completely crossed design was a 2 (criticisms vs. compliment) × 2 (literal vs. ironic

TABLE 1
Examples of Stories

The Weight Lifter Story	
Critical version	
A: I'm good at lifting weights.	
[A fails to lift the weights]	
B: You are bad at lifting weights.	(Literal condition)
or	
You really are good at lifting weights.	(Ironic condition)
Complimentary version	
A: I'm good at lifting weights.	
[A lifts the weights]	
B: You are good at lifting weights.	(Literal condition)
or	
You really are bad at lifting weights.	(Ironic condition)
The Birthday Candles Story	
Critical version	
A: I'm good at blowing out candles.	
[A fails to blow out candles on a cake]	
B: You are bad at blowing out candles.	(Literal condition)
or	
You really are good at blowing out candles.	(Ironic condition)
Complimentary version	
A: I'm good at blowing out candles.	
[A blows out candles on a cake]	
B: You are good at blowing out candles.	(Literal condition)
or	
You really are bad at blowing out candles.	(Ironic condition)

endings) $\times$ 2 (treatment order) $\times$ 2 (5 vs. 6 years of age) mixed factorial with treatment order and age as the only between-subject factors. We note at this point that neither the age factor nor the order of treatment factor accounted for a significant amount of variance in this comprehensive model; consequently, these factors were excluded from our final data analysis strategy.

Procedure.   Participants were tested individually, and all sessions were videotaped. A practice story was employed prior to presenting each set of eight stories to illustrate the procedure. After the child watched each story, three questions were posed. The first question assessed the child's ability to detect the literalness of the speaker's final statement. This first-order belief question was either "Did B really think that A was a good (e.g., basketball player)?" or "Did B really think that A was a bad (e.g., basketball player)?" The order in which the *good* and *bad* queries were presented was randomized for each participant. The second question, "Was B being mean or nice?", assessed the child's comprehension of the speaker's intent. The or-

der in which *mean* and *nice* queries were presented was randomized. Pictorial representations were used to assess the child's comprehension of the speaker's intended meaning. When asked if the final statement in the story was mean or nice, children could respond to the question verbally and by pointing to one of two pictures: one of a happy, nice face and the other of a mean, angry face. For literal statements, the semantic content was congruent with the intended meaning. For nonliteral statements, the semantic content was incongruent with the intended meaning. For example, if an ironic criticism was detected as nonliteral, the second question determined whether the child correctly interpreted a positive statement as negative (mean). Conversely, if an ironic compliment was detected as nonliteral, the second question determined whether the child correctly interpreted a negative statement as positive (nice).

Finally, a third question was asked—either "Do you think that A was a good (e.g., basketball player)?" or "Do you think that A was a bad (e.g., basketball player)?" These questions confirmed whether participants understood the factual context of the story. Again, the order in which the good and bad queries were presented was randomized for each participant.

After children watched the first 8 stories in a randomized order, each participant played a brief game with the experimenter as a distraction while the tapes were changed. The game was a modified version of the Bear-and-Dragon game often used to assess inhibitory control in children (Kochanska, Murray, Jacques, Koenig, & Vandegeest, 1996). After this game, the second set of stories and the previously described procedures were repeated.

Results

No additional analyses were required on results obtained from the third question in the procedure. All participants correctly answered this control question, demonstrating that they were able to represent accurately the factual context of each story (e.g., that a person who scrambles the letters in the alphabet does not know the alphabet). Given the specific questions outlined in the rationale for this study, we proceeded directly to the simple effects tests required to answer these questions (Keppel & Zedick, 1989, pp. 233–234). Because the scores were limited in range (0–4 for each condition), nonparametric, Wilcoxon signed-rank tests were employed for these simple effects tests.

Children's responses to the first question indicated whether they correctly detected the speaker's beliefs about the final utterance in each story (i.e., was the speaker being literal or nonliteral when making this comment?). Table 2 presents these scores converted to proportions for each of the four conditions. Consider first the stories with literal endings. As the data in Table 2 indicate, when asked, children clearly understood that the speaker was being literal when either a literal criti-

TABLE 2
Mean Proportion of Speakers' Beliefs Correctly Detected for Each Story Ending

Criticism				Compliments			
Literal		Ironic		Literal		Ironic	
M	SD	M	SD	M	SD	M	SD
.96	.12	.47	.44	.99	.05	.25	.39

Note. *n* = 24.

cism was employed appropriately to make a negative comment about a poor performance (96%) or a literal compliment was employed appropriately to make a positive comment about a good performance (99%). The near-perfect comprehension of these comments as literal statements confirms (along with results from Question 3) that the children both remembered the factual context and comprehended the sequence of events in these scenarios.

Consider next the stories ending in ironic statements. When the speaker praised a partner's failure (i.e., ironic criticism), the children correctly inferred that the speaker's praise should not be taken literally on 47% of these trials. The appropriate control against which to compare performance in the ironic criticism condition is performance in the literal compliment condition. Children are reacting to the identical statement in both of these conditions (e.g., You sure are a good basketball player). When they encountered this statement in the literal compliment condition, they correctly interpreted the statement as literal on 99% of the trials. In other words, when children encounter the identical statement as an ironic criticism, they are willing to reject their otherwise consistent literal interpretation of the same positive statement and switch to a nonliteral interpretation on approximately half of the trials. A nonparametric Wilcoxon test reveals that scores for the identical statement in the literal praise (99%) and nonliteral ironic criticism conditions (47%) were significantly different ($z = 3.62$, $p < .001$).

A similar pattern of results was observed in the ironic compliment condition. When the speaker criticized a partner's success (i.e., ironic compliment), the children correctly decided that the speaker's criticism should not be taken literally on 25% of the trials. The appropriate control against which to compare performance in the ironic compliment condition is performance in the literal criticism condition. Children are reacting to the identical statement in both of these conditions (e.g., You sure are a bad basketball player). When they encountered this statement in the literal criticism condition, they correctly interpreted the statement as literal on 96% of the trials. In other words, when children encounter the identical statement as an ironic compliment, they are willing to reject a literal interpretation of this same negative statement and switch to a nonliteral interpretation on one fourth of

the trials. A nonparametric Wilcoxon test reveals that their scores for the identical statement in the literal criticism (96%) and the nonliteral ironic compliment conditions (25%) are significantly different ($z = 4.06$, $p < .001$).

In presenting these comparisons, we should perhaps note that an interesting interpretive issue arises with respect to the criterion that should be used to evaluate the children's ability to detect the nonliteral nature of the ironic statements. As Ackerman (1983) pointed out, a 50% chance hypothesis is one possible criterion (e.g., the children either detect or do not detect the nonliteral nature of the final comment). By this criterion, the children detected the nonliteral nature of ironic compliments at levels significantly below chance and detected the nonliteral nature of ironic criticisms at levels roughly equivalent to chance. However, this criterion of performance fails to acknowledge that for most children beyond 3 years of age, the literal form of the comment is perceived as appropriate, relevant, and truthful during most forms of discourse (Dunham & Dunham, 1996; Grice, 1975). Consequently, the more appropriate criterion proposed by Ackerman (1983) and adopted in this study acknowledges this bias to interpret comments literally and measures responses to the ironic statements in terms of their deviation from their literal form as a baseline.

In addition to these comparisons, there are two important points to note in the previously described results. First, there is an asymmetry in the children's performance when detecting the nonliteral nature of the ironic criticisms and compliments. A nonparametric Wilcoxon test confirms this asymmetry indicating that they were more likely to reject a literal interpretation of the ironic criticisms ($z = 2.32$, $p < .02$). This difference in the mean scores is further confirmed by the number of children who managed to detect correctly at least one ironic comment. Fifteen of the children met this criterion for ironic criticisms, whereas only 9 reached this criterion for ironic praise (McNemar test, $p < .03$). Furthermore, no participants met this criterion for detection of ironic praise without also meeting this criterion for detection of ironic criticisms. Considered together, the children's asymmetric performance across the ironic criticism and ironic praise conditions reveals that children are more likely to believe incorrectly that a speaker is being literal with an ironic compliment than with an ironic criticism. The second, more general point is that the absolute levels of detection in both the praise and criticism conditions are well below ceiling levels. Children fail to detect ironic criticisms on 53% of the trials and fail to detect ironic praise on 75% of the trials.

Children's responses to the second question assessed whether they correctly comprehended the speaker's pragmatic intent (i.e., did the speaker intend to be mean or nice when making this comment?). Again, individual scores on this question could range from 0 to 4 in each of the four possible story endings. The number of critical statements rated as mean and complimentary statements rated as nice were converted into proportions and are presented in Table 3. Note that only those participants who detected the nonliteral nature of one or more of both ironic criti-

TABLE 3
Mean Proportion of Correct Comprehension of Speaker's Intended Meaning for
Each Story Ending

Criticisms Rated Mean				Compliments Rated Nice			
Literal		Ironic		Literal		Ironic	
M	SD	M	SD	M	SD	M	SD
.81	.24	.44	.37	1.0	0.0	.58	.46

Note.   These proportions only refer to items in which the speaker's belief was correctly detected. $n = 9$.

cisms and compliments were included in the analysis of pragmatic intent ($n = 9$).
Note also that these proportions refer only to items in which the speaker's
nonliteralness was correctly detected.

Consider first the responses to the mean versus nice question for stories with lit-
eral endings. As the data in Table 3 indicate, when asked, children always (100% of
the time) inferred that the speaker was being nice when a successful performance
ended with a literal compliment. Children inferred that the speaker was being mean
when a poor performance ended with a literal criticism on 81% of the trials.

Turning next to the stories ending in ironic statements, when the speaker
praised a partner's failure (i.e., ironic criticism), the children correctly decided that
the speaker's praise was in fact a criticism and judged it as mean on 44% of these
trials. Again, the appropriate control against which to compare performance in the
ironic criticism condition is performance in the literal compliment condition. The
children are reacting to identical statements in both conditions (e.g., You sure are a
good basketball player). When they encounter this statement in the literal compli-
ment condition, they interpret the speaker's intent as mean on none of the trials. In
other words, when children encounter the same statement as an ironic criticism,
they are willing to reject a literal interpretation of the speaker's intent as nice and
switch to an interpretation of the speaker's intent as mean on approximately half of
the trials. A nonparametric Wilcoxon test reveals that children's ratings of the
same statement as mean in the literal praise (0%) and ironic criticism conditions
(44%) were significantly different ($z = 2.68$, $p < .01$).

A similar pattern of results is observed in the ironic compliment condition. When
the speaker criticized a partner's success (i.e., ironic compliment), the children cor-
rectly decided that the speaker's criticism was in fact complimentary and judged it as
nice on 58% of the trials. The appropriate control against which to compare perfor-
mance in the ironic compliment condition is performance in the literal criticism con-
dition. The children are reacting to identical statements in both conditions (e.g., You
sure are a bad basketball player). When they encounter the same statement in the lit-
eral criticism condition, they interpret the statement as nice on only 19% of the trials.

In other words, when children encounter the same statement as an ironic compliment, they are willing to reject their literal interpretation of the speaker's intent as mean and switch to an interpretation of the speaker's intent as nice on approximately half of the trials. A nonparametric Wilcoxon test reveals that children's ratings of the same statement as nice in the literal criticism (19%) and ironic praise conditions (58%) were marginally different ($z = 1.78$, $p = .075$).

Among the 9 children who had detected that a speaker was being nonliteral in the ironic conditions, the ability to judge correctly a speaker's intent tended to be symmetrical across ironic criticisms and ironic compliments. A comparison of the relevant difference scores revealed that children were as likely to reject a nice interpretation of a positive statement in the ironic criticism condition (0% – 44% = 44%) as they were to reject a mean interpretation of a negative statement in the ironic compliment condition (19% – 58% = 39%). A Wilcoxon test on these difference scores indicated this difference was not significant. Considered together, these data suggest that for this age group, attributions of intent presuppose detection, and the ability to attribute intent once nonliteralness is detected is essentially equivalent for ironic criticisms and ironic compliments.

The preceding analyses of the children's responses to the mean versus nice question included only those children who managed to detect the nonliteral nature of at least one ironic criticism and one ironic compliment in response to the first question. It is also of some interest to look at the responses to the mean versus nice question provided by children who completely failed on the first question to detect the nonliteral nature of the ironic criticisms ($n = 9$) and the ironic compliments ($n = 15$). Essentially, the children who failed to detect any form of irony described 100% of the positive statements (e.g., You sure are a good basketball player) as nice whether that statement was encountered in an ironic or literal context. They described 100% of the negative statements (e.g., You sure are a bad basketball player) as mean whether that statement was encountered in an ironic or a literal context. These data confirm that these particular children, in contrast to those participants who were able to detect nonliteralness, are simply interpreting every statement made during these stories as literal, and they attribute the positive and negative intent of the speaker accordingly on every trial.

Discussion

The objective of this experiment was to determine whether children's emerging abilities to detect the nonliteral nature of ironic statements and to understand pragmatic aspects of irony are symmetrical across ironic criticisms and ironic praise. We should again acknowledge that the children's absolute performance in detecting the nonliteral nature of these ironic comments was not spectacular. They are clearly in the early stages of comprehending that adults do not always mean what

they say. However, in spite of this, the data from the first experiment indicate that rejecting literal interpretations is in fact easier in the context of ironic verbal criticisms than it is in the context of ironic verbal praise. Fifteen of the 24 children detected at least one ironic criticism, whereas only 9 achieved this criterion for ironic compliments, and only those children who detected an ironic criticism were also able to detect ironic compliments. Similarly, these children rejected literal interpretations of ironic statements more frequently when they were intended as critical (44%) than when they were intended as complimentary (25%).

It is also the case that detection of the nonliteral nature of an ironic comment at this age does not guarantee that the correct inference will be made about the speaker's pragmatic intent. These data confirm Ackerman's (1983) observations indicating that detection and intent are separable components of irony comprehension and that children find the latter more difficult than the former. More important, in this study, the asymmetry evident in the detection performance was not observed in the children's comprehension of the speaker's intent to be either mean or nice. Children who correctly detected that the speaker was nonliteral in the ironic conditions were as likely to reject nice interpretations for ironic criticisms (i.e., You sure are a good basketball player) as they were to reject mean interpretations of ironic compliments (i.e., You sure are a bad basketball player). In other words, participants appropriately switched to an interpretation of the speaker's intent as opposite to the semantic content for ironic criticisms and compliments on approximately the same proportion of trials.

Why do these young children find it easier to detect the nonliteral nature of an ironic criticism? The data described earlier by Dews and Winner (1997) indicating that children have more experience with ironic criticism during early development offer one possible explanation for the asymmetry observed in this experiment. This asymmetry would be predicted by theories of early communicative development that place a heavy emphasis on social–cultural learning mechanisms (Bruner, 1983; Tomasello, 1988, 1992, 1995) operating during adult–child social interactions across early development. From a social-learning perspective, the more exposure children have to the various social and linguistic structures that mark verbal irony, the more proficient they should be at detecting nonliteral speech acts and the pragmatic intent of the speaker. As such, the asymmetry can be attributed in a straightforward manner to the differential amount of experience that children have with these two ironic forms during early communicative development. Questions of course remain about the relevant social–pragmatic markers involved in such learning and the optimal social–cultural conditions under which children will acquire these skills.

Although the social learning explanation is intuitively compelling, it is important to note that another factor may also be operating in this situation. In the procedure employed in Experiment 1, there is a potentially important difference in the cues present in the ironic criticism and compliment conditions. Recall that the initial statement in each story was a boast by the first partner (e.g., I'm a *good* basket-

ball player). Although this boast is an identical antecedent statement that sets the context for either an ironic or a literal comment at the end of the story, the speaker's final comment echoes this initial boast *uniquely* in the ironic criticism condition (e.g., You sure are a *good* basketball player). In contrast, the speaker's final statement in the ironic compliment condition (e.g., You sure are a *bad* basketball player) does not echo the initial boast.

This difference is potentially important because studies of irony comprehension in mature conversationalists have suggested that echoic markers may be particularly important in the detection and comprehension of ironic compliments (Kreuz & Glucksberg, 1989; Sperber & Wilson, 1986). Although these theories differ on the exact "nature" of the echoic marker, when Kreuz and Glucksberg presented adults with stories ending with counterfactual negative statements (e.g., the phrase "What awful weather" uttered on a sunny day), these statements were rated as more sarcastic and more sensible when a negative antecedent (e.g., It's probably going to rain tomorrow) was explicitly "echoed" by the ironist. These authors proposed that, for ironic compliments, directly echoing a negative statement facilitates irony comprehension. If, like adults, children at this age are relying on some form of echoic marker to detect ironic compliments, the absence of an explicit negative antecedent in the ironic compliment condition in Experiment 1 may account for the asymmetry we observed in children's ability to detect an ironic compliment as nonliteral.

EXPERIMENT 2

In the second experiment, to test the importance of the previously described echoic factor—the initial boast (e.g., I'm a good basketball player) employed in Experiment 1—stories was replaced with a self-critical statement (e.g., I'm a bad basketball player). This self-critical statement creates an antecedent that will be *uniquely* echoed by an ironic compliment after a successful performance (e.g., You sure are a bad basketball player). If children are using this echoic marker as an important cue for detecting the nonliteral nature of ironic comments, with all other factors equivalent across these two experiments, an interesting reversal in the results should be observed in this second experiment. The addition of the echoic factor to the ironic compliment condition should boost children's ability to detect ironic compliments, and the absence of the echoic factor in the ironic criticism condition should undermine their performance. As such, these results would be the opposite of those observed in Experiment 1.

Method

Participants.   Twenty-four English-speaking, middle-class, 5- to 6-year-olds, again recruited from the database of volunteer families, participated in the study (16 girls, 8 boys). The 12 older children (6 girls, 6 boys) ranged from 76 to

80 months ($M = 77.3$, $SD = 1.2$), and the 12 younger children (10 girls, 2 boys) ranged from 65 to 68 months ($M = 67.0$, $SD = 1.0$). The mean age for all children was 72.2 months ($SD = 5.4$).

Materials.   The nine short stories from Experiment 1 were modified by substituting the initial boast statement made by A with an initial self-critical statement. This was accomplished by dubbing the appropriate critical statement (e.g., I'm a bad basketball player) over the boast (e.g., I'm a good basketball player) for each story. All other aspects of the story (i.e., the action and B's final statement) remained unchanged. Note, however, that in these stories, B's final statement now explicitly echoed the initial statement only in the literal and ironic compliment conditions (e.g., You sure are a bad basketball player).

Design and procedure.   The design was identical to that of Experiment 1. Each participant watched 16 stories: Four ended with a literal criticism, 4 with an ironic criticism, 4 with a literal compliment, and 4 with an ironic compliment. The complete design was a 2 (criticisms vs. compliment) × 2 (literal vs. ironic endings) × 2 (treatment order) × 2 (5 vs. 6 years of age) mixed factorial design with treatment order and age as the only between-subject factors.

The procedure employed in Experiment 1 was replicated as exactly as possible. The stories illustrated in Table 1 were employed again with the previously described change to the initial statement. After the child watched each story, the same three questions described in Experiment 1 were again posed.

Results

As in Experiment 1, initial analysis of the complete model revealed that age and order of treatment factors had no significant effects on any of the dependent measures, and they did not interact with other independent variables. Consequently these variables were dropped from all subsequent analyses. Similarly, no additional

TABLE 4
Mean Proportion of Speakers' Beliefs Correctly Detected for Each Story Ending

Criticisms				Compliments			
Literal		Ironic		Literal		Ironic	
M	SD	M	SD	M	SD	M	SD
.95	.21	.44	.42	1.0	0.0	.35	.38

Note. $n = 24$.

analyses were required on results obtained from the third question in the procedure, as all participants correctly answered this control question regarding the factual context of each story.

Children's responses to the first question indicated whether they correctly detected the speaker's beliefs about the final utterance in each story (see Table 4). Consider first the stories with literal endings. Children clearly understood that the speaker was being literal when either a literal criticism was employed appropriately to make a negative comment about a poor performance (95%) or a literal compliment was employed appropriately to make a positive comment about a good performance (100%).

Consider next the stories ending in ironic statements. When the speaker praised a partner's failure (i.e., ironic criticism), the children correctly inferred that the speaker's praise should not be taken literally on 44% of these trials. A nonparametric Wilcoxon test again revealed that their scores for the identical statement (e.g., You sure are a good basketball player) in the literal praise (100%) and nonliteral ironic criticism conditions (44%) were significantly different ($z = 3.83, p < .001$). In the ironic compliment condition, when the speaker criticized a partner's success (i.e., ironic compliment), the children correctly decided that the speaker's negative comment should not be taken literally on 35% of the trials. A nonparametric Wilcoxon test revealed that their scores for the identical statement (e.g., You sure are a bad basketball player) in the literal criticism (95%) and the nonliteral ironic compliment conditions (35%) were significantly different ($z = 3.47, p < .001$).

The important point to note in these results is that the children's performance when detecting the nonliteral nature of the ironic criticisms and compliments in the second experiment is now essentially symmetrical. A nonparametric Wilcoxon test reveals that the difference in detection rates between the ironic conditions is not significant ($z = 1.03, ns$). The symmetry is confirmed further by the number of children who managed to detect correctly at least one ironic comment. An equal number of children met this criterion for both ironic criticisms ($n = 15$) and ironic praise ($n = 15$). Furthermore, participants were as likely to detect an ironic compliment and no ironic criticisms as they were to detect an ironic criticism and no ironic compliments. Considered together, the children's symmetrical performance across the ironic conditions suggests that children are equally likely to detect either an ironic criticism or an ironic compliment as nonliteral *if* a self-critical antecedent is echoed in the ironic compliment condition.

Children's responses to the second question assessed whether they correctly comprehended the speaker's pragmatic intent. The number of critical statements rated as mean and complimentary statements rated as nice were converted into proportions and are presented in Table 5. Only those participants who detected the nonliteral nature of one or more of both ironic criticisms and compliments were included in the analysis of pragmatic intent ($n = 11$), and the proportions presented refer only to items in which the speaker's nonliteralness was correctly detected.

TABLE 5
Mean Proportion of Correct Comprehension of Speaker's Intended Meaning
for Each Statement Type

Criticisms Rated Mean				Compliments Rated Nice			
Literal		Ironic		Literal		Ironic	
M	SD	M	SD	M	SD	M	SD
.84	.32	.43	.48	1.0	0.0	.46	.47

Note. These proportions only refer to items in which the speaker's belief was correctly detected. $n = 11$.

Consider first the responses to the mean versus nice question for stories with literal endings. As the data in Table 5 indicate, children always (100% of the time) inferred that the speaker was being nice when a successful performance ended with a literal compliment. Children inferred that the speaker was being mean when a poor performance ended with a literal criticism on 84% of the trials. Turning next to the stories ending in ironic statements, when the speaker praised a partner's failure (i.e., ironic criticism), the children correctly decided that the speaker's praise was in fact a criticism and judged it as mean on 43% of these trials. When they encountered this statement (e.g., You sure are a good basketball player) in the literal compliment condition they interpreted the speaker's intent as mean on none of the trials. A nonparametric Wilcoxon test reveals that their ratings of the same statement as mean in the literal praise (0%) and ironic criticism conditions (43%) were significantly different ($z = 2.21$, $p < .05$).

A similar pattern of results is observed in the ironic compliment condition. When the speaker criticized a partner's success (i.e., ironic compliment), the children correctly decided that the speaker's criticism was in fact complimentary and judged it as nice on 46% of the trials. When they encounter this statement (e.g., You sure are a bad basketball player) in the literal criticism condition, they interpret the statement as nice on only 16% of the trials. A nonparametric Wilcoxon test reveals that their ratings of the same statement as nice in the literal criticism (16%) and ironic praise conditions (46%) were significantly different ($z = 2.03$, $p < .05$).

For the 11 children who had detected that a speaker was being nonliteral in both ironic conditions, the ability to judge correctly the speaker's intent tended to be symmetrical across ironic criticisms and ironic compliments. A comparison of the relevant scores revealed that children were as likely to reject a nice interpretation of a positive statement in the ironic criticism condition (0% − 43% = 43%) as they were to reject a mean interpretation of a negative statement in the ironic compliment condition (16% − 46% = 30%). Although this numerical difference continues to favor the ironic criticism condition, a Wilcoxon test on these difference scores indicated this difference was not significant. Considered together, these data suggest again that, for this age group, the ability to attribute intent, once they have de-

tected correctly the nonliteralness of an ironic statement, is essentially equivalent for ironic criticisms and ironic compliments.

The preceding analyses of the children's responses to the mean versus nice question include only those children who managed to detect the nonliteral nature of at least one ironic criticism and one ironic compliment in response to the first question. Responses to the mean versus nice question provided by children who completely failed to detect the nonliteral nature of the ironic criticisms ($n = 9$) and the ironic compliments ($n = 9$) were again of interest. Essentially, the children who failed to detect any form of irony described 100% of the positive statements (e.g., You sure are a good basketball player) as nice whether that statement was encountered in an ironic or literal context. They described 100% of the negative statements (e.g., You sure are a bad basketball player) as mean whether that statement was encountered in an ironic or a literal context. These data confirm that these particular children, in contrast to the participants who were able to detect nonliteralness, are simply interpreting every statement made during these stories as literal, and they attribute positive and negative intent to the speaker accordingly on every trial.

Discussion

As predicted, modifying the stories so that the ironic compliments uniquely echoed a previous negative statement increased the number of children able to detect at least one ironic compliment from 9 children in the first experiment to 15 in the second, whereas the number of children detecting ironic criticisms did not change (Pearson $\chi^2 = 3.00$, one-tailed; $p < .04$). In addition, the echoic cue increased the proportion of ironic compliments detected from 25% in the first experiment to 35% in the second (Mann–Whitney $U = 227$, $p = .09$). Note, however, that removing the echoic factor from the ironic criticism condition in this second experiment did not undermine detection performance. An equal number of children were able to detect at least one ironic criticism in both experiments ($n = 15$), and the proportion of ironic criticisms detected was essentially equivalent across experiments (47% vs. 44%).

Considered together, these data suggest that the echoic factor plays an important, specific role in helping children detect the nonliteral nature of ironic compliments. Indeed, the positive effect of the echoic factor in this second experiment was sufficient to reduce significantly the asymmetry between detection of ironic criticisms and compliments we observed in Experiment 1. The data compared across experiments also indicate, however, that the echoic factor is not particularly important when processing ironic criticisms. Performance did not decline in this second experiment when the explicit echoic factor was removed from the story. Indeed, putting the echoic factor aside, if we simply average detection rates across

the two experiments, ironic criticisms generally continue to be detected more frequently (45% vs. 30%) than ironic compliments ($z = 2.33$, $p < .02$).

The results concerned with the comprehension of the speaker's intent to be either mean or nice confirmed the data from Experiment 1. Again, the intent question was more difficult for children than the detection question, and performance across the ironic conditions was again symmetrical when children did correctly infer intent. Children who detected that the speaker was nonliteral in the ironic conditions were as likely to reject nice interpretations for ironic criticisms (e.g., You sure are a good basketball player) as they were to reject mean interpretations of ironic compliments (e.g., You sure are a bad basketball player).

GENERAL DISCUSSION

The primary question addressed in this study was whether children, during this early transition period of irony comprehension, display a bias in their ability to detect nonliteral forms of criticisms and compliments. When the data concerned with the detection question are considered across both experiments, two conclusions are suggested. First, the superior performance in the ironic criticism condition in Experiment 1 cannot be attributed to the advantage offered by an echoic marker in that condition. Experiment 2 revealed that children performed equally well in detecting ironic criticisms whether the echoic marker was present (Experiment 1) or absent (Experiment 2). Second, the data from Experiment 2 also suggest that the presence of an echoic marker did uniquely and significantly enhance the children's ability to detect the nonliteral nature of ironic compliments.

How does one explain this interaction and the relative advantage enjoyed by the critical ironist in these data? As discussed earlier, one suggestion is that the asymmetry in children's performance reflects their differential experience with this particular form of irony (Dews & Winner, 1997). As such, the results are in general consistent with social–pragmatic theories of early communicative development (e.g., Bruner, 1983; Tomasello, 1988, 1992, 1995) that would attribute these asymmetrical competencies to differential exposure.

Although early social–pragmatic experiences may contribute directly to the differences we have observed, the unique influence of the echoic factor on ironic compliments observed in the second experiment suggests that a simple social-learning mechanism may be an oversimplification. Recent attempts to explain a similar asymmetry in mature conversationalists' comprehension of various forms of verbal irony (Kreuz & Glucksberg, 1989; Kumon-Nakamura et al., 1995) potentially provide a more elaborate explanation for the results of these two experiments. Specifically, Kumon-Nakamura et al. outlined an allusional pretense theory of irony that is also grounded in our social–cultural experiences but explains the asymmetry in terms of the differential extent to which ironic criticisms and

ironic praise echo or allude to some antecedent event, social norm, or shared expectation. Starting from the assumption that our social norms, preferences, and desires tend to be positive (e.g., people desire good weather, expect polite behavior, etc.), allusional pretense theory suggests that ironic criticisms do not require explicit antecedent statements. Instead ironic criticisms *implicitly* echo or allude to our presumed positive expectations (e.g., 'Another gorgeous day' uttered in a downpour). In contrast, ironic compliments (e.g., 'You sure are a terrible friend' after receiving a gift) are assumed to require explicit antecedents (e.g., 'You won't get a gift from a terrible friend like me') because these negative utterances are less effective reminders of our implicit positive expectations or norms.

The data presented in this article suggest that 5- and 6-year-old children are operating in exactly the manner predicted by allusional pretense theory. The explicit echoic statement was influential in the context of the ironic compliments but had no effect in the context of ironic criticisms. More specifically, when a negative antecedent remark was directly echoed in the ironic compliment condition (Experiment 2), the children's detection performance was enhanced and the asymmetry between the two forms of ironic statements was diminished. Equally important, a reciprocal decrease in detection performance was not observed in the ironic criticism condition when the explicit, positive antecedent remark was removed. Apparently, as the allusional pretense model would predict, the positive antecedent condition can be implicitly *assumed* by the critical ironist.

It is also important to emphasize that our extension of allusional pretense theory to these data implies that the wide variety of positive social norms, preferences, and expectations presumed by this model are also in place at this young age. Indeed, it is possible that children's early, differential social–pragmatic experiences with these two forms of verbal irony (Dews & Winner, 1997) are contributing directly to the emergence of the shared expectations that are central to the allusional pretense model. We suspect, however, that some caution is in order with respect to this assumption. Our own view, and something to investigate in further research, is that children may have acquired some positive norms and expectations at this young age, but we would expect this bias to be more domain specific than the general norms assumed for adults.

Finally, we suggest that the allusional pretense model also offers a viable explanation for earlier findings, suggesting that echoic cues do not play an important role in young children's comprehension of irony (e.g., Dews et al., 1996). Recall that all prior research has considered only children's comprehension of ironic criticisms. From the perspective of the allusional pretense model (and consistent with our data), ironic criticisms are assumed to be less dependent on *explicit* antecedents. Consequently, it is perhaps not surprising that echoic cues have failed to play a role in prior developmental studies.

The second question of interest in this article is the degree to which detection and understanding pragmatic intent are separable components of irony compre-

hension during this early period of communicative development. Ackerman (1983) first suggested that the detection of the nonliteral form could be dissociated from the process of inferring the intent or attitude of the speaker. His conclusion was based, in part, on the observation that 6-year-olds were able to detect the nonliteral nature of an ironic criticism more frequently than they were able to infer the negative intent of the ironist. Our results are similar. When responses are averaged across the ironic conditions and experiments, our children were able to detect that a speaker was being nonliteral more frequently (38%) than they were able to correctly infer the speaker's intent (28%), and this difference was significant ($z = 3.02$, $p < .01$).

One would not expect this difference if these two processes were dependent on the same underlying process, and one would not expect the asymmetry in children's performance to occur only in response to the detection question. One admittedly speculative explanation for these results is that our two questions are tapping separable components of irony comprehension that are systematically associated with a fundamental change in social cognition that is also in transition during this period of development. Specifically, a considerable amount of converging evidence suggests that a child's ability to understand the mental states of others is a minimally necessary condition for the comprehension of ironic statements. As Dews and Winner (1997) noted, to understand a nonliteral utterance as ironic, the hearer must make two determinations about the speaker's mental state. First, the hearer must correctly determine the speaker's belief about the situation under discussion. Consider, for example, an ironic criticism ("You sure are a good weight lifter") delivered after a failure to lift the weights. To detect the nonliteral nature of this statement, the listener must correctly infer the speaker's actual belief about the listener's weight-lifting skills. If the listener infers that the speaker does not really believe the positive statement, the listener has inferred the "true" belief of the speaker. This process is indexed by our detection question and is typically described as first-order reasoning about other's belief states. In addition, once this first-order inference is made, to determine the speaker's pragmatic intent or attitude (e.g., is the speaker being mean or nice or possibly lying to the listener), the listener must also infer the *speaker's belief* about the listener's actual knowledge of the situation. Again, in the context of an ironic criticism delivered after a failure to lift the weights, the listener must infer what the speaker believes about the listener's actual knowledge state. In this example, correctly inferring that the speaker believes that the listener knows he or she has failed the weight-lifting task permits correct inferences about the intent of the nonliteral statement (i.e., is the speaker being mean or nice). This process is typically described as second-order reasoning about belief states, and it is indexed by our inference question (Winner & Leekam, 1991).

Several types of evidence also suggest that a child's understanding of irony appears to be constrained by the ability to make these first- and second-order belief attributions. For example, in addition to descriptive data indicating that the ability

to infer second-order belief attributions and inferences about the pragmatic intent of an ironist tend to emerge at about the same age (see Dews & Winner, 1997), more direct research on individual differences has demonstrated that children who fail or pass independent second-order false belief tasks also differ in their ability to comprehend the intent of an ironist (Happe, 1993; Leekam, 1991; Sullivan, Winner, & Hopfield, 1995; Winner & Leekam, 1991).

Given that first-order reasoning skills emerge prior to second-order reasoning skills during early development (e.g., Frye, Zelazo, & Palfai, 1995), we would suggest, consistent with Ackerman's (1983) earlier claims, that the detection of nonliteralness may be a first-order stage of the underlying reasoning process that is separable from and necessary for second-order inferences about the speaker's pragmatic intent. Although this model runs contrary to some current theories of adult irony comprehension (e.g., Gibbs & O'Brien, 1991), it is consistent with the data we have obtained from these young children (see also deGroot, Kaplan, Rosenblatt, Dews, & Winner, 1995). Some of the children in our procedure are clearly able to detect both the nonliteral nature of the ironic comment (first-order attribution) and infer the speaker's intent (second-order attribution). Others, however, detect only the nonliteral nature of the speech and are not consistently able to engage in the second-order reasoning required to infer intent. There are also a substantial number of children at this age who are unable to detect nonliteral speech. These children appear content to take everyone at their word in every context.

Finally, we can conclude this somewhat speculative analysis by returning for a moment to allusional pretense theory and describing its implications in the context of the two-stage process of early irony comprehension previously outlined. Although further research employing a wider age range will be required to address the issue, our data suggest that the mechanisms outlined by allusional pretense theory must be operating specifically on the detection stage of the comprehension process. It is at this stage that we observe the asymmetry in children's comprehension. Once children have detected the nonliteral nature of an ironic compliment or an ironic criticism, if they also possess the second-order belief attribution skills required to infer pragmatic intent, the advantage ascribed to the critical ironist and the role of implicit and explicit echoic antecedents apparently disappears, and children who can manage second-order belief attributions find it equally easy to infer the intent of both forms.

ACKNOWLEDGMENTS

REFERENCES

Ackerman, B. P. (1981). Young children's understanding of a speaker's intentional use of a false utterance. *Developmental Psychology, 17,* 472–480.

Ackerman, B. P. (1983). Form and function in children's understanding of ironic utterances. *Journal of Experimental Child Psychology, 35,* 487–508.

Andrews, J., Rosenblatt, E., Malkus, U., Gardner, H., & Winner, E. (1986). Children's abilities to distinguish metaphoric and ironic utterances from mistakes and lies. *Communication and Cognition, 19,* 281–298.

Bruner, J. (1983). *Child's talk.* New York: Norton.

Clark, H. H., & Gerrig, R. J. (1984). On the pretense theory of irony. *Journal of Experimental Psychology: General, 113,* 121–126.

deGroot, A., Kaplan, J., Rosenblatt, E., Dews, S., & Winner, E. (1995). Understanding vs. discriminating non-literal utterances: Evidence for a dissociation. *Metaphor and Symbolic Activity, 10,* 255–273.

Demorest, A., Meyer, C., Phelps, E., Gardner, H., & Winner, E. (1984). Words speak louder than actions: Understanding deliberately false remarks. *Child Development, 55,* 1527–1534.

Dews, S., & Winner, E. (1997). Attributing meaning to deliberately false utterances: The case of irony. In C. Mandell & A. McCabe (Eds.), *The problem of meaning: Behavioral and cognitive perspectives* (pp. 377–414). Amsterdam: North-Holland/Elsevier Science.

Dews, S., Winner, E., Kaplan, J., Rosenblatt, E., Hunt, M., Lim, K., McGovern, A., Qualter, A., & Marsh, B. (1996). Children's understanding of the meaning and functions of verbal irony. *Child Development, 67,* 3071–3085.

Dunham, P., & Dunham, F. (1996). The semantically reciprocating robot: Adult influences on children's early conversational skills. *Social Development, 5,* 261–274.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10,* 483–527.

Gibbs, R. W. (1994). Figurative thought and figurative language. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 411–446). San Diego, CA: Academic.

Gibbs, R. W., & O'Brien, J. (1991). Psychological aspects of irony understanding. *Journal of Pragmatics, 16,* 523–530.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York: Academic.

Happe, F. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition, 48,* 101–119.

Keppel, G., & Zedick, S. (1989). *Data analysis for research designs.* New York: Freeman.

Kochanska, G., Murray, K., Jacques, T., Koenig, A., & Vandegeest, K. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development, 67,* 490–507.

Kreuz, R. J., & Glucksberg, S. (1989). How to be sarcastic: The echoic reminder theory of verbal irony. *Journal of Experimental Psychology: General, 18,* 374–386.

Kumon-Nakamura, S., Glucksberg, S., & Brown, M. (1995). How about another piece of pie: The allusional pretense theory of discourse irony. *Journal of Experimental Psychology: General, 124,* 3–21.

Leekam, S. (1991). Jokes and lies: Children's understanding of intentional falsehood. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development, and simulation of everyday mindreading* (pp. 159–174). Cambridge, MA: Blackwell.

Milosky, L. M., & Ford, J. A. (1997). The role of prosody in children's inferences of ironic intent. *Discourse Processes, 23,* 47–61.

Sperber, D., & Wilson, D. (1981). Irony and the use-mention distinction. In P. Cole (Ed.), *Radical pragmatics* (pp. 295–318). New York: Academic.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition.* Cambridge, MA: Harvard University Press.

Sullivan, K., Winner, E., & Hopfield, N. (1995). How children tell a lie from a joke: The role of second-order mental state attributions. *British Journal of Developmental Psychology, 13,* 191–204.

Tomasello, M. (1988). The role of joint attentional processes in early language development. *Language Sciences, 10,* 69–98.

Tomasello, M. (1992). The social bases of language acquisition. *Social Development, 1,* 68–87.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Winner, E., & Leekam, S. (1991). Distinguishing irony from deception: Understanding the speaker's second-order intention. *British Journal of Developmental Psychology, 9,* 257–270.