ASYMMETRIES IN THE USE OF VERBAL IRONY

ROGER J. KREUZ
KRISTEN E. LINK
University of Memphis

Three experiments assessed four variables that may affect verbal irony processing: people's expectations of events, event outcome, evaluations of outcome, and shared common ground. Reading times and rating tasks were used to quantify the interaction of these factors. The failed expectation hypothesis predicts an interaction of expectation, outcome, and evaluation. In contrast, the expectation irrelevance hypothesis states that expectation does not matter—only interactions between outcome and evaluation should result. The results provide support for the expectation irrelevance hypothesis. There were also consistent common ground effects: Statements directed at high common ground targets were read more quickly and rated as more ironic than statements directed at low common ground targets. These studies also provide online evidence of the asymmetry of affect (positive evaluations of negative outcomes are more ironic than negative evaluations of positive outcomes). Together, these experiments further elucidate the complex pragmatic factors that govern verbal irony comprehension.

The use of figurative, or nonliteral, language is an important means by which people express their thoughts, feelings, and attitudes to one another (Fussell & Moss, 1998; Ortony, 1975; Roberts & Kreuz, 1994). Metaphor, for example, can be used to express ideas that may be difficult to convey literally (Ortony, 1975). By saying, "My job is a jail," ideas like confinement and lack of control are used to express negative feelings about one's occupation.

Other forms of nonliteral language, such as verbal irony, may be used to express different discourse goals. Verbal irony is typically the expression of an attitude that is in opposition with the true state of affairs (e.g., saying, "What beautiful weather!" during a thunderstorm). The listener must appreciate the discrepancy between what is

---

said and what is meant, and the process by which this occurs has been of interest to researchers in the psychology of language.

Previous research on verbal irony has identified several cues and constraints that play a role in its successful interpretation. Some of these cues involve a particular tone of voice or facial expression (e.g., Cutler, 1974; Kreuz, 1996; Kreuz & Roberts, 1995). Other cues involve the exaggeration of statements to impart ironic intent (Colston & Keller, 1998; Kreuz, Kassler, Coppenrath, & McLain Allen, 1999; Kreuz & Roberts, 1995). Another line of research has examined how echoes of previous assertions affect irony comprehension (Gibbs, 1986, 2000; Jorgensen, Miller, & Sperber, 1984; Kreuz & Glucksberg, 1989). Research on two additional factors, the role of expectations and the influence of common ground, will be reviewed in the following sections, and a series of experiments examining these factors will be reported.

IRONY AND EXPECTATION

One important determinant of the perception of irony is the expectations and assumptions that people bring to a given situation. It has been shown, for example, that situational irony is partially driven by the unexpectedness of particular events (Lucariello, 1994). A similar phenomenon has been noted for verbal irony. It has been suggested that ironic statements are more typically positive evaluations of negative situations (i.e., canonical irony) than negative evaluations of positive situations (i.e., noncanonical irony). For example, it is more common to say, "What lovely weather!" to ironically comment on bad weather than to say, "What miserable weather!" to ironically comment on pleasant weather. This "asymmetry of affect" (Clark & Gerrig, 1984) has been documented in the literature (Kreuz & Glucksberg, 1989; Kumon-Nakamura, Glucksberg, & Brown, 1995; Sperber & Wilson, 1981).

Researchers have tended to explain this asymmetry in terms of a Pollyanna hypothesis (Boucher & Osgood, 1969; Matlin & Stang, 1978)—there are cultural expectations that enterprises will meet with success, that the weather should be pleasant, and so on. When these implicit expectations are violated, statements that echo a positive cultural norm will be readily interpreted as ironic. Negative statements, on the other hand, are not echoes of these expectations and can only be perceived ironically in certain situations, for example, when the statement is an explicit echo of an incorrect prediction (Kreuz & Glucksberg, 1989).

Alternative ways of explaining this asymmetry have been proposed by other researchers. One of these is the negativity effect, which has been studied by Kanouse (1971). It has been shown that people place greater weight on negative information than positive information, all

other things being equal. This may be the result of vigilance (Wegner & Vallacher, 1977)—people are more sensitive to negative information because it may signal dangers or threats. Therefore, people may be more likely to comment on negative events in a distinctive way to signal their importance.

In studying verbal irony, researchers have typically used stimulus materials in which a positive cultural expectation is violated (i.e., canonical irony). However, little is known about how noncanonical irony is processed. In many situations, individuals' expectations may not be positive. For example, visiting a dentist and ending a relationship are situations that most people would not expect to go well or smoothly. Consider the following example:

> *Example 1:* During their patrol of the occupied territory, the GIs discovered an abandoned chateau. They entered cautiously because they were afraid of snipers. It had been raining, and the soldiers were wet and exhausted. However, the chateau had a fully stocked wine cellar, and they began to sample the select vintages. As he put down his glass, the sergeant turned to one of his buddies and said, "War is hell!"

In this situation, the soldier has made a negative evaluation of a positive outcome (finding the wine). However, because the cultural expectations for wartime are negative, the statement can be interpreted ironically. This would be an example of noncanonical irony.

The utterance "War is hell!" is perceived as ironic because of a failed expectation—what is typically expected in war has not happened in this situation (i.e., they find the wine instead). So some negative evaluations do succeed as irony because they highlight the discrepancy between a negative expectation and a positive outcome (see also Colston, 1997). Other negative evaluations, however, will fail as irony because the expectation is not negative. This failed expectation hypothesis states that when expectations are positive, a positive evaluation of a negative outcome (canonical irony) will be processed more readily than a negative evaluation of a positive outcome (noncanonical irony). However, when expectations are negative (as in Example 1), we would expect the opposite: that noncanonical irony will be processed more readily than canonical irony. This would be because the evaluation ("War is hell!") echoes negative expectations that people have about war.[1]

Kumon-Nakamura et al. (1995, Experiment 2) conducted a study in which they presented participants with scenarios utilizing a negative expectation (e.g., the cleanliness of New York subways). For example, the story about the subways was written so that the trains were either clean (positive context) or dirty (negative context). The final statement in each context was either positive ("New York subways are certainly clean!") or negative ("New York subways are certainly dirty!"). There was a significant interaction between context and statement, and

follow-up tests confirmed the prediction that given a negative expectation, a negative statement will be interpreted ironically if it is uttered in a positive context. This is because the statement echoes the implicit negative expectation (i.e., that subways will not be clean) in this context. The authors did not compare the irony ratings of canonical versus noncanonical irony to see if these two forms of irony differ from each other. However, the reported means suggest that canonical irony is perceived as more ironic than noncanonical irony.

However, another alternative is possible. Consider the following example:

> *Example 2:* During their patrol of the occupied territory, the GIs discovered an abandoned chateau. They entered cautiously because they were afraid of snipers. It had been raining, and the soldiers were wet and exhausted. The chateau had a fully stocked wine cellar, and they began to sample the select vintages, but the wine was no good. As he put down his glass, the sergeant turned to one of his buddies and said, "War is wonderful!"

The phrase "War is wonderful!" seems ironic in this context because the outcome is negative. Therefore, any positive evaluation will be perceived nonliterally. This will be referred to as the expectation irrelevance hypothesis: The expectation does not affect irony comprehension, but what does matter is whether the evaluation is inconsistent with the outcome. Canonical and noncanonical irony will both be perceived as ironic, regardless of the expectations of the situation.

Kumon-Nakamura et al.'s (1995) results seem to support the failed expectation hypothesis, but those researchers only evaluated the effect of negative expectations. A true test of the failed expectation hypothesis would include both negative and positive expectations, to see whether expectations about events affect the processing of verbal irony in the same way. The experiments reported here were designed to test these competing hypotheses, using both online and offline measures.

IRONY AND COMMON GROUND

While the Kumon-Nakamura et al. (1995) study is informative, it may well be the case that other factors affect the comprehension of irony when expectations are negative. As Clark and Gerrig (1984) pointed out, "A listener's understanding of an ironic utterance depends crucially on the common ground he or she believes is shared by the ironist and the audience" (p. 124). Clark and his colleagues have repeatedly demonstrated the importance of common ground in language and discourse (e.g., Clark, 1996; Clark & Marshall, 1981). However, common ground has not been manipulated in most previous studies of irony. In the Kumon-Nakamura et al. experiments, for example, ironic statements are uttered by classmates, traveling companions,

and roommates. In all of these cases, high common ground can be presumed to exist between the speaker and the addressee.

Common ground manipulations may not have been investigated before because ironic utterances are typically used in high common ground contexts. Therefore, when researchers write realistic scenarios employing irony, they produce stories that do not vary on this dimension. This may reflect the principle of inferability (Kreuz, 1996), which states that people use irony only when they are reasonably certain that it will be understood as intended. One heuristic that may be employed to determine inferability is the presence of common ground.

A study that examined the role of common ground in irony interpretation was conducted by Kreuz, Kassler, Coppenrath, and McLain Allen (1999). It was found that ironic statements directed at high common ground addressees were judged as more appropriate than those directed at low common ground addressees. However, this study did not manipulate the role of expectation: All of the ironic statements were positive evaluations of negative situations in the context of positive expectations (e.g., going to the movies or dining out).

The three experiments reported here were designed to assess the effects of and interactions among four variables that may affect irony comprehension: whether an expectation is positive or negative, whether the outcome of the event is positive or negative, whether the evaluation of the outcome is positive or negative, and the degree of common ground shared by listener and speaker.

In all of the experiments reported here, the reading time for the potentially ironic remark will serve as a measure of comprehension time. The three experiments differ, however, in the ratings provided by participants. In Experiment 1, participants were asked to rate the perceived irony of these remarks. Experiment 2 was designed to assess the semantic aspects of irony: Participants provided ratings of how much sense these remarks made. The third experiment, in contrast, was designed to assess pragmatic factors. Participants provided ratings of how appropriate these remarks seemed to be.

This design allowed us to test the following predictions:

1. Canonical irony will be processed as irony more readily than noncanonical irony, demonstrating the asymmetry of affect in both reading times and ratings.
2. The failed expectation hypothesis predicts a three-way interaction of expectation, outcome, and evaluation. Alternatively, the expectation irrelevance hypothesis predicts a two-way interaction of outcome and evaluation, but not the three-way interaction.
3. The existence of high common ground between speaker and addressee, in comparison to low common ground, will facilitate the processing of ironic statements.

As we have seen, although researchers have examined the role of expectation and common ground separately, these factors have not

been combined into one study. Such a combination would be desirable because it would allow for the examination of higher order interactions among these factors. This would, for example, allow for a determination of how common ground interacts with other cues, such as the outcome of an event.

EXPERIMENT 1

METHOD

Participants. Thirty-two undergraduates at the University of Memphis participated for course credit. All of the participants indicated that they were native speakers of English.

Materials. A pilot study was conducted to assess the events to be used in the experimental scenarios. The experimenters generated 30 positive and 30 negative events. These included events in which a positive experience would be expected (e.g., winning a lottery, getting a raise), or a negative experience would be expected (e.g., getting mugged, finding hair in one's food). These 60 events were presented to 29 participants, who rated them on a 6-point scale, ranging from *very positive* to *very negative*. We selected the 16 events rated most positively and the 16 rated most negatively to be the topics of our experimental scenarios. The mean for the positive events was 1.47 (SD = .28), and the mean for the negative events was 5.63 (SD = .19).

Five-sentence scenarios were built around each event from the pilot study. In the first sentence of each scenario, the principal characters were introduced. In addition, an event, which could lead to a positive or negative expectation, was described. In the second and third sentences, the situation was elaborated further. In the fourth sentence, the outcome of the situation was revealed. For example, in the hair-in-food story, either the characters get their meal for free (positive outcome) or the waiter is indifferent to the discovery (negative outcome). In the fifth sentence, an evaluation was made by one of the characters, which was either positive (e.g., "Aren't you glad we came to such a classy place?") or negative (e.g., "Aren't you glad we came to such a low-class place?"). In addition, this evaluation was spoken to an addressee who shared high or low common ground with the speaker. For example, the high common ground addressee in the hair-in-food story is the speaker's mother, and the low common ground addressee in that story is a person at a neighboring table in the restaurant. To use Clark's (1996, p. 115) terms, the pairs sharing high common ground were "intimates" or "friends," and the pairs sharing low common ground were "acquaintances" or "strangers."

Table 1
Example of a Positive Expectation Scenario Used in the Experiments

"The Beauty Pageant"

  Sentences 1 to 3
    Tiffany couldn't believe it, but she had won the title of Miss Tennessee at the
      state beauty pageant.
    She had dieted for months and spent thousands of dollars on her dress, but it
      had all been worth it.
    The talent portion had gone well, and she placed first in the evening gown
      competition.
  Sentence 4—Outcome manipulation
    Positive outcome
      As Tiffany was crowned and began her walk down the runway, she smiled
        radiantly at the TV cameras and the judges.
    Negative outcome
      As Tiffany was crowned and began her walk down the runway, the heel on her
        shoe broke, and she fell off the runway onto the judges.
  Sentence 5—Common ground and evaluation manipulation
    High common ground, positive evaluation
      Crying, Tiffany said to her parents, "Tonight has been just perfect!"
    High common ground, negative evaluation
      Crying, Tiffany said to her parents, "Tonight has been just awful!"
    Low common ground, positive evaluation
      Crying, Tiffany said to one of the judges, "Tonight has been just perfect!"
    Low common ground, negative evaluation
      Crying, Tiffany said to one of the judges, "Tonight has been just awful!"

Sixteen of the scenarios were constructed around the positive events from the pilot study, and 16 were built around the negative events from that study. Each scenario could appear in one of eight versions (i.e., two types of outcomes × two types of evaluations × two levels of common ground), for a total of 256 individual scenarios. Examples of these scenarios appear in Tables 1 and 2.

Design and procedure. Participants were tested individually at Macintosh workstations. The presentation of instructions and materials was controlled by PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993). Eight different lists of scenarios were employed. Each list contained one version of all 32 scenarios. Within each list, there was a random assignment of version to list, so that no participant ever saw more than one version of each scenario. A repeated measures design was employed, in which each participant saw two scenarios with the same combination of variables (e.g., positive expectation, negative outcome, negative evaluation, and high common ground). Each list was randomly assigned to each participant, and each participant saw a randomized order of the scenarios in his or her list.

At the beginning of each session, the participants read a series of instructions on the computer screen. Participants were told that they

Table 2
Example of a Negative Expectation Scenario Used in the Experiments

"The Robbery"

Sentences 1 to 3
  John and his wife, Lisa, came home from a movie one night and realized that
    their door had been forced open.
  Inside, they found that most of their belongings were missing.
  Lisa was devastated by the loss of her grandmother's silver.

Sentence 4—Outcome manipulation
  Positive outcome
    While John talked to the police officers, another officer drove up with the
      moving van that contained all of their things.
  Negative outcome
    While John talked to the police officers, another officer drove up and remarked
      that the situation seemed hopeless.

Sentence 5—Common ground and evaluation manipulation
  High common ground, positive evaluation
    John said to Lisa, "This is really great news."
  High common ground, negative evaluation
    John said to Lisa, "This is really rotten news."
  Low common ground, positive evaluation
    John said to the officer in the van, "This is really great news."
  Low common ground, negative evaluation
    John said to the officer in the van, "This is really rotten news."

would read a series of scenarios and provide ratings of sarcasm or irony. These terms were not defined for the participants, but they were provided with a brief example.

Each scenario was presented one sentence at a time. Participants read each sentence at their own pace and pressed a key when they were ready to read the next sentence. The new sentence replaced the old sentence at the top of the computer screen. At the end of each scenario, the participants were presented with a 6-point scale and asked, "How ironic or sarcastic was the final sentence?" The endpoints of the scale were labeled *not at all* and *very*. The participants indicated their rating by typing a numeric response from 1 through 6. At this point, they pressed a key to see the next scenario. Participants continued in this way until they had seen all 32 scenarios. The experimental program recorded the reading time for each sentence (i.e., the time between each key press). In addition, the numeric rating was also recorded.

RESULTS

To test the hypotheses, we performed the following planned comparisons for both dependent variables: (a) differences between canonical and noncanonical irony; (b) three-way interactions between expectation, outcome, and evaluation, or two-way interactions between

Table 3
Interaction of Outcome by Evaluation in the Reading Time Data From Experiment 1

	Type of Outcome	
Type of Evaluation	Positive	Negative
Positive	4,494 (1,452)	4,711 (1,552)
Negative	5,438 (1,892)	4,810 (1,683)

Note. Standard deviations are in parentheses.

outcome and evaluation; and (c) the main effects of common ground and its interactions with outcome and evaluation.

Reading times. An inspection of the participants' reading times for the final sentence of each scenario was conducted to detect the presence of outliers. Values that were more than 2.5 standard deviations above and below each participant's mean were replaced by that participant's mean. This resulted in the replacement of 2.5% of the values. These reading times were analyzed using a 2 (positive vs. negative expectation) × 2 (positive vs. negative outcome) × 2 (high vs. low common ground) × 2 (positive vs. negative evaluation) repeated measures analysis of variance (ANOVA). Following Clark (1973), analyses were conducted using both participants as the random effect (F_1) and materials as the random effect (F_2). (When materials were treated as the random effect, expectation was a between-subjects factor, and the other variables were treated as within-subjects factors.)

There was a main effect of common ground, $F_1 (1, 31) = 4.17, p = .05, F_2(1, 30) = 2.42, p = .13$. Sentences in which the addressee shared high common ground with the speaker ($M = 4,735$ ms, $SD = 1,509$) were read more quickly than sentences with low common ground addressees ($M = 4,991$ ms, $SD = 1,639$). However, the nonliteral statement reading times did not differ in terms of common ground in the participant analysis, $F_1(1, 31) = 1.78, p = .19$; this result was, however, significant in the item analysis in the predicted direction, $F_2(1, 31) = 5.08, p < .05$.

There was an interaction of outcome and evaluation; $F_1(1, 31) = 8.74, p < .01; F_2(1, 30) = 8.36, p < .01$; see Table 3. The literal statements (i.e., positive evaluations/positive outcomes and negative evaluations/negative outcomes) were read more quickly ($M = 4,652$ ms, $SD = 1,509$) than the nonliteral statements (i.e., negative evaluations/positive outcomes, and positive evaluations/negative outcomes) ($M = 5,074$ ms, $SD = 1,662$), and this difference was significant; $F_1(1, 31) = 8.74, p < .01; F_2(1, 31) = 8.51, p < .01$.

A comparison of the canonical ($M = 4,711$ ms) and the noncanonical ($M = 5,438$ ms) ironic statements revealed a significant difference; $F_1(1, 31) = 18.22, p < .01; F_2(1, 31) = 13.59, p < .01$. This finding is consistent with the asymmetry of affect described earlier. Post hoc analyses indicated that canonical irony was read as quickly as the literal

Table 4

Interaction of Outcome, Evaluation, and Common Ground in the Irony Rating Data From Experiment 1

	Type of Outcome	
Type of Evaluation	Positive	Negative
Positive		
High common ground	2.14 (1.06)	4.84 (1.06)
Low common ground	2.38 (1.11)	4.58 (1.24)
Negative		
High common ground	4.58 (1.22)	3.06 (1.20)
Low common ground	4.09 (1.31)	3.05 (1.22)

Note. Standard deviations are in parentheses.

statements (4,711 ms vs. 4,652 ms, Fs < 1), which is concordant with Gibbs's (1986, Experiment 1) finding that ironic statements can be processed even more quickly than literal statements. In addition, noncanonical irony was read more slowly than either the literal statements, $F_1(1, 31) = 21.13$, $p < .01$; $F_2(1, 31) = 18.03$, $p < .01$, or canonical irony, $F(1, 31) = 18.22$, $p < .01$; $F_2(1, 31) = 13.59$, $p < .01$.

However, the interaction of expectation, outcome, and evaluation predicted by the failed expectation hypothesis was nonsignificant, Fs < 1. In other words, the asymmetry of affect is not affected by whether background expectations are positive or negative, and this supports the expectation irrelevance hypothesis.

Irony ratings. The ratings for each scenario were also analyzed with a $2 \times 2 \times 2 \times 2$ ANOVA. There was an interaction of outcome and evaluation, but these factors were part of a three-way interaction with common ground; $F_1(1, 31) = 6.82$, $p < .05$; $F_2(1, 30) = 3.83$, $p = .06$. These means appear in Table 4.

As expected, the literal statements were rated as having little irony ($M = 2.66$) compared to the nonliteral statements ($M = 4.52$), and this difference was significant; $F_1(1, 31) = 64.20$, $p < .01$; $F_2(1, 31) = 124.80$, $p < .01$. There is also evidence for the asymmetry of affect in these data: The canonical irony ratings ($M = 4.71$) differed significantly from the noncanonical ratings ($M = 4.34$) in the participant analysis, $F_1(1, 31) = 4.51$, $p < .05$, but not in the item analysis, $F_2(1, 31) = 2.85$, $p = .10$. Finally, these results are affected by the common ground manipulation: Irony ratings for the nonliteral statements are significantly higher when the characters share high common ground ($M = 4.71$) than when the characters share low common ground ($M = 4.33$); $F_1(1, 31) = 10.31$, $p < .01$; $F_2(1, 31) = 10.10$, $p < .01$. Once again, however, the interaction of expectation, outcome, and evaluation was nonsignificant; $F_1(1, 31) = 2.78$, $p = .11$; $F_2(1, 31) = 0.94$, $p = .34$.

EXPERIMENT 2

Experiment 2 differed from Experiment 1 only in the rating that the participants provided. Specifically, they were asked to judge how much sense the potentially ironic remarks made. It was expected that ironic statements would be perceived as making less sense than literal statements, based on previous research examining similar figures of speech, such as hyperbole (Kreuz, Kassler, & Coppenrath, 1998). Moreover, it was predicted that canonical irony would make more sense than noncanonical irony. Finally, the design allowed us to determine whether expectation interacted with outcome and evaluation and allowed a test between the failed expectation and the expectation irrelevance hypotheses.

METHOD

Participants. Thirty-two undergraduates at the University of Memphis participated for course credit. All of the participants indicated that they were native speakers of English, and none had participated in Experiment 1.

Materials. The same scenarios used in Experiment 1 were used in Experiment 2.

Design and procedure. The only difference in the procedure between Experiment 1 and Experiment 2 concerned one of the dependent measures. In this study, participants were asked, "How much sense did the final sentence make?" The endpoints of the 6-point scale were labeled *no sense* and *a lot* (of sense).

RESULTS

Reading times. Reading time outliers for the final sentence were replaced by participants' means, resulting in the replacement of 1.9% of the values. As in Experiment 1, the results were analyzed with $2 \times 2 \times 2 \times 2$ ANOVAs, with both participant and item analyses. An ANOVA of the final sentence reading time revealed a main effect of common ground, $F_1(1, 31) = 6.82, p < .05; F_2(1, 30) = 7.01, p < .05$. An examination of the nonliteral statements revealed a trend in the predicted direction: Reading times were faster when the characters shared high common ground ($M = 4,446$ ms, $SD = 1,405$) than when the characters did not share high common ground ($M = 4,772$ ms, $SD = 1,564$), $F_1(1, 31) = 3.17$, $p < .09$. This effect was significant in the item analysis, $F_2(1, 31) = 5.27$, $p < .05$.

Table 5
Interaction of Outcome by Evaluation in the Sense Rating Data From Experiment 2

	Type of Outcome	
Type of Evaluation	Positive	Negative
Positive	4.75 (0.74)	2.38 (0.77)
Negative	1.95 (0.66)	4.02 (0.85)

Note. Standard deviations are in parentheses.

There was no difference in reading times for the literal statements ($M = 4,625$ ms, $SD = 1,196$) compared to the nonliteral statements ($M = 4,609$, $SD = 1,394$), Fs < 1.

In addition, canonical irony was not read more quickly ($M = 4,501$ ms, $SD = 1,372$) than noncanonical irony ($M = 4,717$ ms, $SD = 1,548$); $F_1(1, 31) = 1.88, p = .18; F_2(1, 31) = 2.48, p = .13$. As in the previous analyses, the interaction of expectation, outcome, and evaluation was nonsignificant, Fs < 1. Once again, this result fails to support the failed expectation hypothesis.

Sense ratings. There was a main effect of evaluation; $F_1(1, 31) = 29.35, p < .01; F_2(1, 30) = 11.41, p < .01$, but this factor was involved in a two-way interaction with outcome; $F_1(1, 31) = 183.36, p < .01; F_2(1, 30) = 171.08, p < .01$. This result provides support for the expectation irrelevance hypothesis. The cell means appear in Table 5.

As would be expected, the literal statements were rated as making more sense ($M = 4.38$) than the nonliteral statements ($M = 2.16$); $F_1(1, 31) = 183.36, p < .01; F_2(1, 31) = 171.20, p < .01$. Canonical irony was rated as making more sense ($M = 2.38$) than noncanonical irony ($M = 1.95$), $F_1(1, 31) = 15.55, p < .01$; this result was marginal in the item analysis, $F_2(1, 31) = 3.92, p = .06$. Once again, the interaction of expectation, outcome, and evaluation was nonsignificant; $F_1(1, 31) = 1.30, p = .26; F_2 = 0.98, p = .33$.

EXPERIMENT 3

The only difference between Experiment 3 and Experiments 1 and 2 was the rating that the participants provided. Specifically, they were asked to judge how appropriate the potentially ironic remarks were. Once again, it was expected that ironic statements would be perceived as being less appropriate than literal statements (Kreuz et al., 1998). It was predicted that canonical irony would be perceived as more appropriate than noncanonical irony. Finally, the results will be examined for the presence of higher order interactions involving expectation, which will distinguish between the failed expectation and the expectation irrelevance hypotheses.

METHOD

Participants. Thirty-two undergraduates at the University of Memphis participated for course credit. All of the participants indicated that they were native speakers of English, and none had participated in Experiments 1 or 2.

Materials. The same scenarios used in Experiments 1 and 2 were used in Experiment 3.

Design and procedure. The only difference in the procedure between Experiment 2 and the present experiment concerned one of the dependent measures. In this study, participants were asked, "How appropriate was the final sentence?" The endpoints of the 6-point scale were labeled *not at all* and *very*.

RESULTS

Reading times. Reading time outliers for the final sentence were replaced by participants' means, resulting in the replacement of 2.3% of the values. As in Experiments 1 and 2, the results were analyzed with $2 \times 2 \times 2 \times 2$ ANOVAs, with both participant and item analyses. There was a main effect of common ground, $F_1(1, 31) = 14.86, p < .01$, $F_2(1, 30) = 8.37, p < .01$. As predicted, nonliteral sentences with a high common ground addressee were read more quickly ($M = 4,626$ ms, $SD = 1,465$) than nonliteral sentences with a low common ground addressee ($M = 5,083$ ms, $SD = 1,384$); $F_1(1, 31) = 5.52, p < .05; F_2(1, 31) = 6.32, p < .05$.

The literal statements were not read more quickly ($M = 4,952$ ms, $SD = 1,508$) than the nonliteral statements ($M = 4,855$ ms, $SD = 1,314$), Fs < 1. In addition, canonical irony was not read more quickly ($M = 4,790$ ms) than noncanonical irony ($M = 4,919$ ms), Fs < 1. As before, the interaction of expectation, outcome, and evaluation was nonsignificant, Fs < 1, which is contradictory to the predictions made by the failed expectation hypothesis.

Appropriateness ratings. An analysis of the ratings revealed a main effect of evaluation, but this factor interacted with outcome; $F_1(1, 31) = 155.38, p < .01; F_2(1, 30) = 224.01, p < .01$. The cell means appear in Table 6. Participants rated the literal statements as being more appropriate ($M = 4.32, SD = .85$) than the nonliteral statements ($M = 2.11, SD = .65$); $F_1(1, 31) = 155.38, p < .01; F_2(1, 31) = 224.37, p < .01$. Canonical irony was rated as more appropriate ($M = 2.29$) than noncanonical irony ($M = 1.92$) in the participant analysis, $F_1(1, 31) = 10.65, p < .01$; this effect was marginal in the item analysis, $F_2(1, 31) = 3.12, p = .09$. Finally, the interaction of expectation, outcome, and evaluation was nonsignificant; $F_1(1, 31) = 2.00, p = .17; F_2 < 1$.

Table 6
Interaction of Outcome by Evaluation in the Appropriateness Rating Data From Experiment 3

Type of Evaluation	Type of Outcome	
	Positive	Negative
Positive	4.68 (0.84)	2.29 (0.81)
Negative	1.92 (0.64)	3.97 (1.02)

Note. Standard deviations are in parentheses.

GENERAL DISCUSSION

The results of these experiments provide a consistent picture with regard to the hypotheses described earlier. The reading time data from Experiment 1 and the rating data from all three studies indicate that canonical irony is read more quickly, is more ironic, makes more sense, and is regarded as more appropriate than noncanonical irony. The rating data from Experiment 1 replicate the rating data from Kumon-Nakamura et al. (1995). However, this study also demonstrates a processing time difference. These data add to a growing literature that demonstrates how the use of verbal irony is constrained by situational, social, and cognitive factors.

The reading times for canonical and noncanonical irony in Experiments 2 and 3, however, did not differ, and this result may have to do with the rating tasks themselves. When participants were explicitly asked about irony, their reading times showed a sensitivity to the manipulations that was not observed when they were asked about other dimensions of these statements.

The failed expectation hypothesis was not supported in these studies. The effect predicted by this hypothesis—an interaction of expectation, outcome, and evaluation—was not observed for any of the dependent measures in any of the experiments. In contrast, there is some support for the expectation irrelevance hypothesis: The predicted interaction of outcome and evaluation was found in the reading time and irony rating data in Experiment 1 and the sense and appropriateness rating data in Experiments 2 and 3. These results are also consistent with Kumon-Nakamura et al. (1995, Experiment 2) in that, for situations with negative expectations, both canonical and noncanonical irony were rated similarly. However, the current research has shown that the same results do not hold for positive expectations. One might conclude that allusions to expectations do not play a major role in the processing of verbal irony.

The common ground manipulation, in contrast, had a very consistent effect on participants' reading times: In all three experiments, participants read statements directed at high common ground

addressees more quickly than those directed at low common ground addressees. This further supports the principle of inferability (Kreuz, 1996), which would predict that people could process irony more easily when it is directed by a speaker to a high common ground addressee. Furthermore, the experiments reported above demonstrate that comprehension time is facilitated by the presence of high common ground.

In addition, the irony ratings were higher in the high common ground condition than in the low common ground condition. Previous research indicates that individuals use verbal irony more with high common ground addressees than with low common ground addressees (Kreuz, 1996). Therefore, when participants were confronted with scenarios in which irony was directed at a low common ground addressee, they may have been uncertain how to interpret such a remark.

Despite the fact that participants were functioning as "overhearers," which might be expected to weaken the effect, common ground effects were still obtained. The robust common ground effects in these studies, therefore, underscore just how important this factor may be in irony comprehension.

It is interesting to note that significant differences exist even for the literal statements. In Experiment 1, negative evaluations of negative outcomes ("negative literals") were perceived as more ironic than positive evaluations of positive outcomes ("positive literals"), $F(1, 31) = 21.05$, $p < .01$. Experiment 2 revealed that negative literals were perceived as making less sense than positive literals, $F(1, 31) = 19.30$, $p < .01$. Finally, the data from Experiment 3 suggest that negative literals are seen as less appropriate than positive literals, $F(1, 31) = 26.11$, $p < .01$.

Why might this be? Because ironic statements typically express negative affect (Roberts & Kreuz, 1994), any negative statement, even a literal one, may have an ironic undertone. However, an inspection of the means reveals that the negative literal statements ($M = 3.05$) were still below the midpoint of the irony rating scale (3.5). In the case of the sense and appropriateness measures, the negative literal statements were above the scale's midpoint (3.94 and 3.97, respectively). Therefore, the negative literals appear to have an ironic tinge, and this result may simply mirror the fact that ironic statements are perceived as making less sense and being less appropriate. Since most of the final statements in the experiments reported here were self-referential, it may also be that it is less face saving to make a negative remark ("This is really rotten news" in Table 2) than a positive remark ("This is really great news").

A number of conclusions can be drawn from these experiments. First, robust common ground effects were observed in all three of the experiments, demonstrating the importance of this factor in verbal irony comprehension. Second, evidence has been presented for the asymmetry of affect, both in online and offline tasks: Canonical irony is

perceived differently than noncanonical irony, regardless of whether the expectation is positive or negative. Finally, we have demonstrated that expectations about events do not influence the interpretation of ironic statements. These findings do not support a failed expectation view of verbal irony but largely support the expectation irrelevance hypothesis.

Clearly, there is a complex interplay between social and cognitive factors in the comprehension of verbal irony. Cultural expectations and negativity effects may play a role with regard to issues of asymmetry, and an awareness of discrepancies between expectations and outcomes is the result of cognitive processing. Other factors, however, can only be described in terms of social and cognitive phenomena. The accumulation of common ground requires a long-term representation of shared knowledge and experiences, and without this, social interaction would be impossible. The importance of such factors becomes clear when complex pragmatic phenomena, such as verbal irony, are considered.

NOTE

1. It is interesting to note that in Example 1, the soldier could also have said, "War is wonderful!" and the comment still would be perceived as an appropriate one, because it still highlights the discrepancy between negative expectation and positive outcome.

REFERENCES

Boucher, J., & Osgood, C. E. (1969). The Pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behavior*, *8*, 1-8.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335-359.

Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.

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

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). Cambridge, UK : Cambridge University Press.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments & Computers*, *25*, 257-271.

Colston, H. L. (1997). "I've never seen anything like it": Overstatement, understatement, and irony. *Metaphor and Symbol*, *12*, 43-58.

Colston, H. L., & Keller, S. B. (1998). You'll never believe this: Irony and hyperbole in expressing surprise. *Journal of Psycholinguistic Research*, *27*, 499-513.

Cutler, A. (1974). On saying what you mean without meaning what you say. In *Papers from the tenth regional meeting of the Chicago Linguistics Society* (pp. 1-11). Chicago: University of Chicago, Department of Linguistics.

Fussell, S. R., & Moss, M. M. (1998). Figurative language in emotional communication. In S. R. Fussell & R. J. Kreuz (Eds.), *Social and cognitive approaches to interpersonal communication* (pp. 113-141). Mahwah, NJ: Lawrence Erlbaum.

Gibbs, R. W., Jr. (1986). On the psycholinguistics of sarcasm. *Journal of Experimental Psychology: General*, *115*, 3-15.

Gibbs, R. W., Jr. (2000). Irony in talk among friends. *Metaphor and Symbol*, *15*, 5-27.

Jorgensen, J., Miller, G. A., & Sperber, D. (1984). Test of the mention theory of irony. *Journal of Experimental Psychology: General*, *113*, 112-120.

Kanouse, D. E. (1971). *Language, labeling and attribution*. Morristown, NJ: General Learning Press.

Kreuz, R. J. (1996). The use of verbal irony: Cues and constraints. In J. S. Mio & A. N. Katz (Eds.), *Metaphor: Implications and applications* (pp. 23-38). Mahwah, NJ: Lawrence Erlbaum.

Kreuz, R. J., & Glucksberg, S. (1989). How to be sarcastic: The echoic reminder theory of verbal irony. *Journal of Experimental Psychology: General*, *118*, 374-386.

Kreuz, R. J., Kassler, M. A., & Coppenrath, L. (1998). The use of exaggeration in discourse: Cognitive and social facets. In S. R. Fussell & R. J. Kreuz (Eds.), *Social and cognitive approaches to interpersonal communication* (pp. 91-111). Mahwah, NJ: Lawrence Erlbaum.

Kreuz, R. J., Kassler, M. A., Coppenrath, L., & McLain Allen, B. (1999). Tag questions and common ground effects in the perception of verbal irony. *Journal of Pragmatics*, *31*, 1685-1700.

Kreuz, R. J., & Roberts, R. M. (1995). Two cues for verbal irony: Hyperbole and the ironic tone of voice. *Metaphor and Symbolic Activity*, *10*, 21-31.

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

Lucariello, J. (1994). Situational irony: A concept of events gone awry. *Journal of Experimental Psychology: General*, *123*, 129-145.

Matlin, M. W., & Stang, D. J. (1978). *The Pollyanna principle: Selectivity in language, memory, and thought*. Cambridge, MA: Schenkman.

Ortony, A. (1975). Why metaphors are necessary and not just nice. *Educational Theory*, *25*, 45-53.

Roberts, R. M., & Kreuz, R. J. (1994). Why do people use figurative language? *Psychological Science*, *5*, 159-163.

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

Wegner, D. M., & Vallacher, R. R. (1977). *Implicit psychology: An introduction to social cognition*. New York: Oxford University Press.