UNIVERSITY OF CALGARY

Processing Irony: Ratings, Reading Time, and Priming

by

Stacey Ivanko

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF PSYCHOLOGY

CALGARY, ALBERTA

June 26, 2001

Abstract

There are two prominent theories of how irony is processed: (1) irony is processed by activating the ironic meaning directly (e.g., Gibbs, 1994); and (2) irony is processed by first examining and rejecting the literal meaning (e.g., Giora, Fein, & Schwartz, 1998). The present research examined the role of context (using strongly negative, weakly negative, and neutral contexts) in processing irony. In the strongly negative context, for ironic statements, reading times were longer than for literal statements, and priming data showed that ironic and literal meanings were both activated at an interstimulus interval (ISI) of 100 ms. In the weakly negative context, for ironic statements, reading times were the same as for literal statements, and only the ironic meaning was activated at an ISI of 100 ms. These results showed that ironic meanings can be activated directly, and that context plays an important role in the processing of ironic statements.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Processing Irony: Ratings, Reading Time, and Priming

Irony[1] is a common form of figurative language. An ironic statement is generally identified as such because of an incongruity that is present between the statement and the preceding context. In the case of ironic criticism, the context would be negative and the surface meaning of the statement would be positive. This incongruity would lead the reader/listener to interpret the statement as ironic and understand the negative underlying meaning.

Figurative language has many uses, one of which seems to be a way of communicating ideas, attitudes and thoughts in a manner that does not explicitly state them, especially if these attitudes and thoughts are negative. Even though these ideas and attitudes are not stated explicitly, we can understand them with relative ease. Our ability to use figurative language so effortlessly is related to the fact that much of our "thinking is constrained by figurative processes" (Gibbs, 1994, p. 413). The ease with which we understand figurative language is also a function of common beliefs and knowledge that are shared between the speaker and listener. For instance, if the speaker and listener are unfamiliar with each other, the listener may have a more difficult time determining if the speaker's words are literal or figurative.

There are at least two compelling reasons to study how irony is processed. The

---

[1] I will be using the term "irony" to refer to the form of verbal irony which is commonly referred to as "sarcasm." The items used in the current experiments have a statement which has an underlying meaning that is approximately opposite to the surface (literal) meaning. This is most common of sarcasm and is the form of irony which is most typically examined.

first reason is that irony is used frequently in every day language and it is understood with relative ease. Irony appears to serve several functions in communication: we use irony to mock, to be funny, to criticize (Kreuz, Long, & Church, 1991), to lighten a situation (Dews & Winner, 1995), to a convey certain attitude, and to create a certain social impression (Winner & Gardner, 1993). It is an important type of communication, and yet we have very few theories that describe how irony is understood and we have even fewer empirical studies examining how irony is processed. Thus, one reason for studying irony is to correct this and contribute to a coherent psychological theory of irony.

The second reason for studying irony is that it is one type of figurative language that individuals use. By understanding how irony is processed, this will help us to understand how other types of figurative language are interpreted. Irony is one of many types of ambiguity in language. As such, by learning how the ambiguities are resolved (e.g., the cues that are used, and the processes that are used for disambiguation), we will likely understand the cues used to resolve other types of ambiguity in language.

The purpose of the current study is to investigate how we process irony. Some researchers suggest that we can process irony without activating and assessing the literal interpretation of the statement first (e.g., Gibbs, 1986; 1993; 1994; Kreuz & Glucksberg, 1989), but other researchers have suggested that the literal interpretation of the statement must be examined and rejected before we can consider the ironic meaning (e.g., Giora, Fein, & Schwartz, 1998; Giora, 1995). The goal of the present research was to test these theories by comparing processing time for literal and ironic interpretations. Very few previous studies have made this comparison.

We use many cues to help us understand that a speaker is being ironic and to determine what the speaker intends in the ironic statement. Such cues include tone of voice, facial cues, and most important, context. I refer to context as being the most important cue because we are able to understand irony without tone of voice or facial cues as we come across irony in novels. Without any context about the situation being referred to, it would be very difficult (or impossible in the absence of all cues) to determine that a statement is ironic. If there is nothing to suggest otherwise, a statement is believed to be literal if it appears without any type of cue indicating that it is ironic. Gibbs (1994) argues that "the ease with which many figurative utterances are produced and comprehended is due in part to the context for linguistic understanding, or more specifically, common ground (i.e., the knowledge, beliefs, and attitudes that are recognized as being shared by speakers and listeners in any discourse situation)" (p. 413). If, for example, someone said that painting is very beautiful, without any context to accompany the statement, it is very difficult to know if it is ironic or literal. However, if the listener knows that the speaker does not like any work produced by the artist because they are abstract and the painting being referred to in the statement is abstract, then the listener can be rather certain (without any other cues) that the speaker is being ironic.

Research has shown that irony is perceived as a result of the incongruity between the context and the statement (e.g., Gerrig & Goldvarg, 2000; Katz & Lee, 1993; Katz & Pexman, 1997; Kreuz & Glucksberg, 1989; Pexman, Ferretti, & Katz, 2000; Pexman & Olineck, in press). Gerrig and Goldvarg examined the effect of the degree of situational disparity (e.g., 5 minutes late vs 50 minutes late) on the perception of irony. They found

that greater situational disparity leads to a higher perception of irony than lower situational disparity. Thus, there is clearly an effect of degree of disparity on the perception of irony. It seemed possible that the degree of negativity in the context might also affect processing of ironic criticisms. Failure to control degree of negativity of context in previous on-line studies (e.g., Gibbs, 1986; Giora et al., 1998) might explain the different findings observed for processing ironic criticisms. In the current study I investigated this possibility by manipulating the context preceding the target statement.

Equivalent Processing for the Literal and Figurative Meanings

Gibbs (1986; 1993; 1994) argued that it takes no longer to process the figurative meaning of a statement than it does to process the literal meaning of a statement. This is made possible by the contextual information that is available. Figurative language can be processed as such when given the appropriate contextual information, so that if a mother who has a child with a messy room says, I love children who keep their rooms clean, this would be interpreted as meaning the woman's child has not cleaned his or her room. However, this statement would be less clear if the listener did not know the child has a messy room.

Gibbs (1994) stated that "recognition of the incongruity between what people say and what they do reflects the cognitive ability of people to think ironically" (p. 437). Thus, understanding irony requires no special cognitive processes, and is not particularly effortful. Instead, understanding irony requires parallel activation of literal and figurative meanings. This seems evident since, in some situations, while we are speaking ironically, we may also be making a literal statement. For example, a driver may say "I love people

who signal" after being cut off by another driver. While the speaker is being ironic in the sense that the other driver did not use his signal, this statement also reflects the speaker's literal belief.

Gibbs (1986) examined the processing time involved in reading and understanding ironic, nonironic, literal and acknowledgment statements and the time involved in determining if a paraphrase statement matched the meaning of the last statement read. There was a negative context condition (e.g., Harry was building an addition to his house. He was working real hard putting in the foundation. His younger brother was supposed to help. But he never showed up. At the end of a long day, Harry's brother finally appeared. Harry was a bit upset with him. Harry said to his brother.) where the ironic statements (which were positive in surface meaning; e.g., you're a big help) and nonironic statements (which were negative; e.g., you're not helping me) were used. There was also a positive context condition (e.g., Greg was having trouble with calculus. He had a big exam coming up and he was in trouble. Fortunately, his roommate tutored him on some of the basics. When they were done, Greg felt he'd learned a lot. "Well," he said to his roommate.) where the ironic statements were used literally (which were positive; e.g., you're a big help) and an acknowledgment statement (e.g., thanks for your help) was used.

Gibbs (1986) presented participants with a sentence on a computer screen. They read the sentence and then pressed a button on the keyboard when they understood what that sentence meant, to advance to the next sentence. Following participants' response to the last (target) sentence, a paraphrase of the target sentence appeared (e.g., for the ironic

target you're a fine friend, the paraphrase sentence was you're a bad friend). They then made a true/false judgement for the paraphrase sentence (whether it meant the same as the target sentence or not) by pressing a designated key on the keyboard.

Gibbs (1986) found that the ironic statements were processed faster than the nonironic statements but equivalently to the literal statements. Also, the paraphrase judgements for ironic statements were faster than those for the nonironic statements or for the literal statements. As processing times for the ironic statements were faster than processing times for nonironic statements, this suggested that a literal meaning does not have to be processed first. The judgement times also lend evidence to support this conclusion.

Dews and Winner (1995) provided some criticisms for the methodology used by Gibbs (1986). They argued that the on-line reading measures used may not fully capture the processing involved in comprehending irony. Full comprehension of irony may not occur until after the initial comprehension of the statement, and, therefore, the reading time measurements may not capture this processing time involved in irony comprehension. Dews and Winner also argued that the measures may not be sensitive enough to capture the differences that may occur. They argued that if the literal meaning is processed but ceases to be processed once the statement is recognized to be ironic, the reading time measures may not capture these differences as the reading time of entire statement is measured. In the present research, I improved on these methods by measuring the time taken to read each word in the target statement and by adding a sentence after the target statement to capture any processing that occurs after the statement has been read.

McElree and Nordlie (1999) have examined reading times for metaphors. They presented participants with metaphors, one word at time. Each word was on the screen for 250 ms. After the final word in the metaphor, there was a response lag (from 28 ms to 2500 ms) and then a tone was presented, signaling participants to respond as to whether the string was meaningful or not. There were three types of strings: (1) literal (e.g., some temples are stone); (2) figurative (e.g., some hearts are stone); and (3) nonsense (e.g., some clouds are stone).

McElree and Nordlie (1999) found no evidence that figurative statements take longer to interpret than literal statements, even though the figurative strings were less likely to be judged meaningful. Some criticisms of this paper can be made. First, it is not known whether the participants in the study actually understood the metaphors presented. That is, participants were not asked about their interpretations of the metaphors used, so we do not know whether participants were actually interpreting the metaphors in a figurative manner. A second criticism is the fact that the participants in this study rated the metaphors as less meaningful. Typically, however, we use figurative language to convey more meaning than literal language. This also gives evidence for the first criticism. If figurative language typically conveys more meaning, then it is possible that the stimuli used were not completely understood metaphorically since they were rated as less meaningful. In the present study I will try to avoid these problems by obtaining ratings data for the target stimuli to ensure that the target statements are interpreted in the intended way (ironic or literal).

Processing of the Literal Meaning Before the Figurative Meaning

The second view of ironic processing is that the literal interpretation of the statement must be processed first, before the ironic interpretation can be processed (Giora, 1995). Giora argued that irony does not involve a cancellation of the indirectly negated message, "rather, it entertains both the explicit and implicated messages so that the dissimilarity between them may be computed" (p. 240). Therefore, irony involves what is said (the literal meaning) and what is implicated (the ironic meaning). Giora concluded that an ironic statement is more difficult to process than a literal statement and will therefore take longer to process.

Giora (1995) reinterpreted Gibbs' (1986) findings. Gibbs had taken those findings to mean that irony does not take longer to process than literal language. Giora examined the texts that were used and argued, however, that the ironic targets were more appropriate than the nonironic targets because the ironic targets provided more information. For example, the ironic target, you're a big help, provides more information than the nonironic target, you're not helping me, in the following text:

> Harry was building an addition to his house. He was working real hard putting in the foundation. His younger brother was supposed to help. But he never showed up. At the end of a long day, Harry's brother finally appeared. Harry said to his brother:

Giora argued that the ironic target should take less time to read because it is more informative, not because, as Gibbs argued, irony is not more difficult to understand than literal language.

Giora (1995) also argued that the ironic target should have been compared to the same statement in a literal text to determine ease of processing. This comparison was made and it showed that the ironic utterance took longer to read than the same statement uttered literally. Giora argued that this is evidence that irony is more difficult to process than literal language. One problem with Giora's analysis is that it involved a comparison of reading times for statements that followed either a negative context or a positive context. There is the possibility, however, that negative contexts are generally read more slowly than positive contexts. Pexman and Olineck (2001) have provided evidence to this effect. They examined processing of positive and negative metaphors in irony biasing (negative) and metaphor biasing (positive) contexts, using reading time data. Their results showed that the negative contexts were, in general, read slower than the positive contexts, regardless of the statement used. Thus, comparing the same statement across different contexts can lead to differing reading times due to the context, not the statement.

Giora et al. (1998) proposed a graded salience hypothesis of irony. This hypothesis predicts that a meaning with higher salience, which is typically the literal meaning of the statement, should be activated first. Giora et al. argued that the more salient meaning is interpreted first, which happens to be the literal meaning and not the figurative meaning, and then the literal and figurative meanings are considered together. According to this hypothesis, it would take longer to read figurative statements than literal statements. They presented participants with paragraphs that were constructed of three or four sentences of context and a final statement. The final statement was either literal (a positive statement following a positive context; e.g., Anna is a great student and

very responsible. One day she called to tell me she did not know when she would be able to show up for my lecture. However, just as I was starting, she entered the classroom. I said to her: "You are just in time") or sarcastic (a positive statement following a negative context; e.g., Anna is a great student, but she is very absent-minded. One day while I was well through my lecture, she suddenly showed up in the classroom. I said to her: "You are just in time"). The text was presented to participants line by line and they pressed a button on the keyboard when they finished reading each line and answered a yes/no comprehension question at the end of each paragraph. Consistent with the graded salience hypothesis, Giora et al. found that the figurative statements took longer to read than the literal statements.

Giora et al. (1998) also examined lexical decision response times to words presented just after the target statement in Experiment 2. These words were either related to the literal meaning or to the figurative meaning of the statements. The target sentences (e.g., you are just in time) were presented in a figurative or a literal context, and the word following the sentences supported the literal meaning (e.g., punctual), the figurative meaning (e.g., late) or was a nonword. The words were presented at an interstimulus interval (ISI) of 150 ms or 1000 ms.

In both ISI conditions, Giora et al. (1998, Experiment 2) found that the fastest response times (RTs) were for the test words that were literally related to the statements, whether it was in a literal or figurative context. When an ISI of 2000 ms was used (Experiment 3), RTs were equivalent for literally related and figuratively related words. This supports the hypothesis that the more salient (literal) meaning is activated first,

before the less salient (figurative) meaning is activated. According to these results, it is not until 2000 ms that the figurative meaning is activated.

Giora and Fein (1999) examined familiar and less familiar irony using a lexical decision task (LDT) similar to that used by Giora et al. (1998). To determine familiar and less familiar ironies, participants wrote down a meaning or meanings for the statements which were presented without any preceding context. Statements receiving an ironic interpretation by more than half of the participants were classified as familiar irony and statements receiving an ironic interpretation less than half of the time were classified as less familiar irony.

For the less familiar ironic statements, Giora and Fein (1999) found that the literally related test words were responded to faster than the figuratively related test words in the literal context and in the figurative context. For the familiar ironic statements, however, they found no differences between the literal and figurative test words, suggesting that for familiar irony, the ironic meaning is salient, supporting the graded salience hypothesis.

Thus, while some research suggests equivalence for processing the literal and figurative meanings (e.g., Gibbs, 1986), other research suggests primacy for the literal meaning in processing (e.g., Giora et al., 1998), except for highly familiar ironies (Giora & Fein, 1999). This processing issue has become a central one (yet unresolved) in the irony literature (e.g., Pexman, Ferretti, & Katz, 2000; Toplak & Katz, 2000). In order to attempt to resolve this controversy, the present study examined the role of context in the processing of figurative (ironic) and literal language. I hypothesized that context strength

might influence processing times. With a very negative context, participants might be more likely to expect irony, and their expectations might facilitate subsequent processing of an ironic remark. With a less negative context, participants might not expect an ironic remark. Equivalent processing times for ironic and literal statements might only be observed in a strongly biasing (strongly negative) preceding context. A more weakly biasing (weakly negative) preceding context might lead to a processing advantage for literal over ironic statements. Support for this possibility can be found in the metaphor literature.

In on-line studies of metaphor, results have shown that when the context preceding the metaphor is relatively long, metaphors are processed in the same time as their literal interpretations, but when the context is relatively short, metaphors take longer to process (e.g., Shinjo & Meyers, 1987; Gerrig & Healy, 1983; Ortony, Schallert, Reynolds & Antos, 1978). Ortony et al. presented participants with a short context (3 to 11 words) or a long context (33 to 60 words) followed by a statement that was either literal or metaphorical, depending on the context bias. The paragraphs were presented line by line on a computer, and participants pressed the spacebar on the keyboard to advance each line. They found that target statements were read more quickly when they were presented after the long context than when they were presented after the short context. In the long context, there was no difference in processing time between the literal and metaphorical statements, but in the short context, the literal statements were processed faster than the metaphorical statements.

Gerrig and Healy (1983) have shown that the placement of the metaphor in the

context also affects the processing of that metaphor. They placed the statements either before the context or after the context. They found that metaphors were processed in the same time as a literal interpretation when the context came before the metaphor, but if the context appeared after the metaphor, metaphors were processed more slowly than literal statements.

Since most of the on-line literature involves metaphor, and the focus of the present study is irony, the reader should note that while irony and metaphor are both considered to be examples of figurative language, it is not clear whether metaphors are processed in the same manner as irony. Pexman et al. (2000) conducted an on-line study of metaphor and irony, which provides some indirect evidence for this idea. All the statements used by Pexman et al. (2000) were metaphors which were placed in a metaphor-inducing context or an irony-inducing context. For example, the statement children are precious gems, can be built into a metaphor-inducing context, which would carry the meaning that children are valuable. The same statement can also be used in an ironic sense, so that if a child is misbehaving and this statement is uttered, then the interpretation would be that children are invaluable or troublesome. The results showed a difference in processing time for metaphors and ironic statements. This difference appeared in the space after the target statement, whereby metaphors in the irony-inducing context had a longer reading time at this space than metaphors in the metaphor-inducing context. While the metaphoric content may be available early, the ironic content required additional time to process. This suggested that while metaphor and irony are both forms of figurative language, the way in which they are processed may be different.

Winner and Gardner (1993) argued that irony and metaphor, while they have a common structure in that what the speaker says is not what the speaker means, are different forms of figurative language and may be processed in different ways. They argued that "metaphor interpretation is constrained only by the listener's domain knowledge. ... In contrast, ... irony interpretation is made possible by the ability to infer other people's beliefs, and their beliefs about beliefs" (p. 427). Winner and Gardner distinguished two levels of understanding for metaphor and irony: interpretation and metalinguistic awareness. For interpretation to be reached, the listener must understand what the speaker means. A metalinguistic awareness is reached when the listener recognizes that there is a difference between what is said and what is meant.

Winner and Gardner (1993) suggested that if there are differences in processing for metaphor and irony, it would be evident in the age at which children understand these forms of language. If understanding develops at the same age for both metaphor and irony, they are likely not processed differently from each other. If understanding develops at different ages, this would suggest that the processes involved are different. Research shows that children are able to understand metaphor at a very young age, as long as they have the domain knowledge required for the metaphor (Keil, 1986). For example, her cheeks were roses, could not be understood by a child if s/he does not know that roses are red. Keil demonstrated that children do understand metaphors when they have knowledge of the referents in the metaphor. Irony, however, is not understood by children until around 6 years of age (e.g., Demorest, Meyer, Phelps, Gardner & Winner, 1984; Winner, Windmueller, Rosenblatt, Bosco, Best & Gardner, 1987). Children need to understand

beliefs about another person's belief state (Winner & Gardner, 1992); for example, the recognition that, when Joe knows Kim has a bad haircut, Joe believes that Kim does not know she has a bad haircut. Thus, if Joe said to Kim, you have a great haircut, without a developed theory of mind, children will not understand that Joe is speaking ironically and means that Kim does not have a great haircut. Understanding of irony and metaphor develop at different ages. Metalinguistic understanding, however, develops at the same age for metaphor and irony, at around 6 or 7 years of age (Winner & Gardner, 1993). This demonstrates that irony and metaphor differ in some respects, but they also have similarities. It is, therefore, possible that these two forms of figurative language are processed in different ways.

The Present Research

The present study used three context types which differed only in degree of negativity: strongly negative, weakly negative, and neutral. These contexts were paired with ironic and literal statements. To ensure equivalence in plausibility and familiarity and differences in negativity, the stimuli were pilot-tested in Experiment 1. The stimuli that met criteria for inclusion were then used in four subsequent experiments measuring interpretation and processing.

The stimuli were rated by participants for sarcasm, mocking, and politeness, in Experiment 2, to determine whether the ironic statements were really interpreted ironically. In Experiment 3a I examined the word-by-word reading times for ironic and literal statements. The stimuli were placed in a moving windows paradigm which measured the time taken to read each word in each paragraph presented. Reading times

for words in the target statement were examined as well as reading times for words in a wrap-up sentence following the statement. Reading times for the wrap-up sentences were examined to assess processing that occurred after the statement was read. It was hypothesized that ironic statements in the strongly negative context would be processed in the same time or more quickly than the literal statements, but that in the weakly negative context, the literal statements would be processed faster than the ironic statements. Experiment 3b examined participants' expectations. Participants were presented with the context paragraphs and were asked which statement they expected after that situation.

Experiment 4a was an LDT for probe words that were used in Experiment 4b. This LDT was conducted to determine if there were any baseline response differences for the three categories of probe words. The mean RTs from the LDT were then used to calculate effect sizes (response time for probe word - baseline response time for word in LDT) for the responses in Experiment 4b.

In Experiment 4b, the stimuli were again presented in a moving windows paradigm, but each word was presented for a set duration of 250 ms (thus, it was not a self-paced task). After the target statement, a probe word that related to the literal meaning of the statement, the ironic meaning of the statement, a neutral word, or a nonword, was presented. These words were presented after an ISI of 100 ms or 1000 ms. It was hypothesized that for the strongly negative context, there would be evidence of facilitation for the ironically related words, compared to the literal words, whereas for the weakly negative context, there would be evidence of facilitation for the literal words

compared to the ironic words.

Experiment 1 - Pilot-test

Purpose

The purpose of this experiment was to pilot-test a large number of potential statements and contexts in order to determine ratings of familiarity, plausibility and positivity/negativity for the statements and plausibility and positivity/negativity for the contexts. The aim was to choose 12 pairs of statements and 12 sets of contexts. Each statement pair has a literal (negative) and ironic (positive) version, in which the statements in each pair were matched for familiarity, plausibility, and number of words. Each context set has a strongly negative context, a weakly negative context, and a neutral context, in which the contexts were matched for plausibility.

Method

Participants

There were 65 (54 female, 11 male) undergraduate students from the University of Calgary, aged 17 to 46 ($M = 23.95$, $SD = 6.81$), who participated in this study. Participation was voluntary and participants received bonus credit towards a psychology course in exchange for participation.

Materials

There were 29 situations created, each with four statements and five or six versions of the contexts. In total, there were 116 statements and 149 contexts created (see the Appendix for examples). Two different lists of stimuli were used, so that each participant received 58 statements (one positive and one negative) and 87 contexts (three

from each situation), which were randomly presented in two different booklets. After each statement, 7-point rating scales for familiarity (1 = not at all familiar and 7 = very familiar), plausibility (1 = not at all plausible and 7 = very plausible) and positivity/negativity (1 = very negative and 7 = very positive) were presented. Presented after each context were 7-point rating scales for plausibility and positivity/negativity.

Procedure

Participants were tested in small groups of 5 to 10 individuals. They were randomly assigned to the lists and half of the participants completed the statement booklet first while the other half completed the context booklet first.

Results and Discussion

The mean ratings for each statement and context scenario were examined for each of the 29 situations. Twelve stimuli sets were required to be chosen for later experiments. The goal was to choose sets that had statement pairs matched for familiarity and plausibility, where one statement was positive while the other was negative. For the contexts, the goal was to choose sets that were matched for plausibility, where one context in each set was strongly negative, one weakly negative, and one neutral.

For the 12 statement pairs, the positive statements were to be the ironic statements (e.g., Jason is a good driver) in later experiments and the negative statements were to be the literal statements (e.g., Jason is a poor driver). Each statement was five words in length. Across the statement pairs, the positive statements were significantly more positive ($M = 6.13$, $SD = .28$) than the negative statements ($M = 2.13$, $SD = .42$), $t(11) = 26.29$, $p < .001$, $SE = .15$. The positive and negative statements in the pairs were equally

familiar (positive: M = 5.59, SD = .38; negative: M = 5.35, SD = .51), t(11) = 1.98, p > .05, SE = .12, and were equally plausible (positive: M = 5.81, SD = .22; negative: M = 5.66, SD = .43), t(11) = 1.45, p > .05, SE = .10.

Each context was three sentences long. Within a context set, wording of the contexts (strong, weak, neutral) was identical, except in the second sentence. For example:

> Jason was driving April home from school. Jason cut off another driver and barely avoided a collision / Jason drove very quickly and made tight turns / In the car, Jason and April made plans for the weekend. The next day April is explaining to Lynn what happened. April says:

The change made for the strong, weak, and neutral contexts, respectively, appears in italics (see the Appendix for more examples).

For the context sets, the strongly negative context (M = 2.06, SD = .32) was significantly more negative than the weakly negative context (M = 3.07, SD = .31), t(11) = 14.53, p < .001, SE = .07, which, in turn, was significantly more negative than the neutral context (M = 4.86, SD = .41), t(11) = 16.02, p < .001, SE = .11. The strongly negative context (M = 5.38, SD = .52) was equally as plausible as the weakly negative context (M = 5.55, SD = .31), t(11) = 1.14, p > .05, SE = .15, and the neutral context (M = 5.66, SD = .20), t(11) = 1.91, p > .05 SE = .15. The weakly negative context was also equally as plausible as the neutral context, t(11) = 1.45, p > .05, SE = .07. Bonferroni adjustments were used to control for alpha inflation across all comparisons.

Experiment 2

Purpose

This experiment was designed to determine whether participants understood the ironic intent of the statements that were chosen from Experiment 1. This was to ensure that the statements that were meant to be ironic were being interpreted as ironic. A second purpose was to examine memory performance for literal and ironic statements. Memory performance tends to be facilitated when statements are processed extensively (e.g., Gibbs, 1987; O'Brien and Myers, 1985). Thus, memory scores can be taken as rough indicators of processing. If, as Giora (1995; Giora, et al., 1998) suggested, ironic statements are more difficult to process than literal statements, one might expect better memory for ironic statements (reflecting more extensive processing for those statements). This possibility was tested in this experiment. A third purpose of this experiment was to determine probe words (to be used in Experiment 4) for each statement that supported the literal meaning of each statement and probe words that supported the ironic meaning for each statement.

Method

Participants

There were 89 (65 female, and 24 male) undergraduate students from the University of Calgary, aged 18 to 44 ($M = 21.66$, $SD = 4.66$), who participated in this study. Participation was voluntary and participants received bonus credit towards a psychology course in exchange for participation. There were six versions of the stimuli and 15 participants were assigned to each, with one version having only 14 participants:

one participant did not follow instructions and was excluded from the analysis.

Materials

Twelve sets of statements and contexts derived from Experiment 1 were used, involving a 3 (Context: strongly negative context, weakly negative context, neutral context) by 2 (Statement: ironic statement, literal statement) combination of statements and contexts. Each statement appeared in each context condition across the six versions of the stimuli.

The stimuli were presented in booklets. After each context-statement paragraph were five lines for the probe words and four 7-point scales: (1) sarcasm (1 = not at all sarcastic and 7 = very sarcastic); (2) mocking (1 = not at all mocking and 7 = very mocking); (3) politeness (1 = not at all polite and 7 = very polite); and (4) confidence of rating (1 = not at all confident and 7 = very confident).

There was a distractor task involving nine multiplication and subtraction problems. There was also a free recall task in which participants were asked to recall as many of the target statements (which were presented in bold) as they could.

Procedure

Participants were tested in small groups of 5 to 10 individuals. They completed the ratings booklet first in which they were asked to write down the first five words that came to mind immediately after reading the paragraph and then to complete each rating. After the ratings were complete, 5 minutes was spent on the distractor task and then the free recall task was completed.

Results and Discussion

The mean ratings and memory scores for each combination of statement and context were calculated (see Table 1) and were analyzed by subjects (F_1 and t_1) and by items (F_2 and t_2) using separate 3 (Context: strongly negative, weakly negative, and neutral) x 2 (Statement: sarcastic and literal) repeated-measures analyses of variance (ANOVAs). We also conducted planned comparisons (which were simple main effects for the interaction) for statement effects within each context condition.

Correlations between the five ratings dimensions and the memory scores were also calculated for the strongly negative and weakly negative contexts together, and for the neutral context alone (see Table 2). From the results below, it is clear that the statements in the strongly negative and weakly negative contexts were interpreted in a similar way. However, the statements in the neutral context appeared to be interpreted differently. Sarcasm ratings and mocking ratings in the strongly negative and weakly negative contexts were significantly correlated, $r(697) = .53$, $p < .001$, such that when sarcasm ratings were higher, mocking ratings were also higher. This correlation was also present in the neutral context, $r(355) = .47$, $p < .001$. This correlation supports the idea that one of the functions of irony is to be mocking (Kreuz, et al., 1991). Politeness ratings were significantly correlated with mocking ratings in the strongly negative and weakly negative contexts, $r(697) = -.36$, $p < .001$, as well as the neutral context, $r(355) = -.69$, $p < .001$. Statements rated high on mocking tended to be rated as less polite than statements rated low on mocking.

Sarcasm Ratings

For the sarcasm ratings, there was a significant Context x Statement interaction ($F_1(1, 88) = 29.07$, $p < .001$, $MSE = 1.55$; $F_2(1, 11) = 15.14$, $p < .01$, $MSE = .36$). The nature of this interaction was revealed with planned comparisons. The comparison of statements across the strongly negative context showed that the ironic statements were rated significantly more sarcastic than the literal statements ($t_1(88) = 23.19$, $p < .001$, $SE = .17$; $t_2(11) = 17.14$, $p < .001$, $SE = .22$). The comparison of statements across the weakly negative context also showed that the ironic statements were rated significantly more sarcastic than the literal statements ($t_1(88) = 13.55$, $p < .001$, $SE = .34$; $t_2(11) = 7.73$, $p < .001$, $SE = .34$). The comparison of statements across the neutral context, however, showed the opposite effect: the literal statements were rated significantly more sarcastic than the ironic statements ($t_1(88) = 5.15$, $p < .001$, $SE = .19$; $t_2(11) = 3.16$, $p < .01$, $SE = .31$). The opposite results in the neutral context can be explained by the somewhat positive ratings these contexts received in the pilot study ($M = 4.86$, $SD = .41$). Because the "literal" statements are negative, these statements in the neutral context were likely interpreted as ironic compliments, as indicated by the sarcasm ratings. In the following example, only forgetting two lines in a 20 line poem after studying it for 5 minutes is rather positive.

> Meredith told Kevin that she could memorize a 20 line poem in 5 minutes. Meredith recited the poem and only forgot 2 lines. The next day Kevin is talking to Harry about the poem. Kevin says: Meredith has a weak memory.

Saying that Meredith has a weak memory can be interpreted as an ironic compliment, to

mean that Meredith has an amazing memory.

The main effect of Context was significant ($F_1(1, 88) = 22.66$, $p < .001$, $MSE = 1.02$; $F_2(1, 11) = 15.62$, $p < .01$, $MSE = .17$). The sarcasm ratings for statements presented in the strongly negative context and the weakly negative context ($M = 3.86$, $SD = 2.36$; $M = 3.51$, $SD = 2.14$, respectively) were generally higher than statements presented in the neutral context ($M = 2.35$, $SD = 1.71$). The main effect of Statement was also significant ($F_1(1, 88) = 261.52$, $p < .001$, $MSE = 1.76$; $F_2(1, 11) = 96.47$, $p < .001$, $MSE = .62$). The sarcasm ratings for the ironic statements ($M = 4.11$, $SD = 2.33$) were significantly higher than those for the literal statements ($M = 2.40$, $SD = 1.65$).

These results showed that the statements that were meant to be sarcastic in the strongly negative and weakly negative contexts were, as expected, interpreted as sarcastic.

Mocking Ratings

For the mocking ratings, there was a significant Context x Statement interaction ($F_1(1, 88) = 21.74$, $p < .001$, $MSE = 1.16$; $F_2(1, 11) = 14.00$, $p < .01$, $MSE = .22$). The nature of this interaction was revealed with planned comparisons. The comparison of statements across the strongly negative context showed that the ironic statements were rated significantly more mocking than the literal statements ($t_1(88) = 6.92$, $p < .001$, $SE = .22$; $t_2(11) = 7.04$, $p < .001$, $SE = .21$). The comparison of statements across the weakly negative context also showed that the ironic statements were rated significantly more mocking than the literal statements ($t_1(88) = 2.51$, $p = .01$, $SE = .19$; $t_2(11) = 1.76$, $p > .05$, $SE = .25$). The comparison of statements across the neutral context showed the opposite effect, with the literal statements rated significantly more mocking than the

ironic statements ($t_1(88) = 12.65$, $p < .001$, $SE = .19$; $t_2(11) = 11.16$, $p < .001$, $SE = .21$). This indicated that statements considered to be sarcastic are more mocking than statements that are considered literal in the strongly negative and weakly negative contexts, as indicated by the correlation between sarcasm and mocking. In the neutral context, the "literal" statement was rated as sarcastic, and therefore was perceived to be more mocking than the positive "ironic" statement. This was because, as mentioned above, the neutral contexts were actually quite positive.

The main effect of Context was significant ($F_1(1, 88) = 19.96$, $p < .001$, $MSE = 1.59$; $F_2(1, 11) = 19.44$, $p = .001$, $MSE = .20$). The mocking ratings for statements presented in the strongly negative contexts and the weakly negative contexts ($M = 4.46$, $SD = 1.88$; $M = 4.21$, $SD = 1.83$, respectively) were generally higher than statements presented in the neutral context ($M = 2.97$, $SD = 1.97$). The main effect of Statement was not significant ($F_1 < 1$; $F_2(1, 11) = 1.28$, $p > .05$, $MSE = .34$).

Politeness Ratings

For the politeness ratings, there was a significant Context x Statement interaction ($F_1(1, 88) = 12.91$, $p < .001$, $MSE = .98$; $F_2(1, 11) = 15.47$, $p < .01$, $MSE = .10$). The nature of this interaction was revealed with planned comparisons. The comparison of statements across the strongly negative context showed that the ironic statements were rated significantly more polite than the literal statements ($t_1(88) = 2.55$, $p = .01$, $SE = .21$; $t_2(11) = 2.02$, $p = .07$, $SE = .25$). The comparison of statements across the weakly negative context also showed that the ironic statements were rated significantly more polite than the literal statements ($t_1(88) = 7.71$, $p < .001$, $SE = .18$; $t_2(11) = 4.47$, $p = .001$,

SE = .32). Dews and Winner (1995) proposed the tinge hypothesis, suggesting that one purpose for irony is to mute negative attitudes or beliefs in a situation. Using an ironic criticism is therefore more polite than a literal criticism because the positive surface meaning tinges the negative ironic meaning with positivity. The comparison of statements across the neutral context also shows that the ironic statements were rated more polite than the literal statements ($t_1(88) = 27.40$, $p < .001$, SE = .13; $t_2(11) = 16.92$, $p < .001$, SE = .21). The politeness rating did not show the reverse effect for the neutral context as did the sarcasm and mocking ratings. This is because ironic compliments are considered to be less polite than literal compliments (Pexman & Olineck, 2001). An ironic compliment is one where a negative statement is made in a positive situation. Therefore, when something negative is stated, even though it is a "compliment," it is less polite than a direct compliment, and may be used to indicate a negative attitude towards the victim of the irony.

The main effect of Context was significant ($F_1(1, 88) = 7.89$, $p < .01$, MSE = 1.28; $F_2(1, 11) = 4.42$, $p = .06$, MSE = .32). The politeness ratings for the statements presented in the neutral context (M = 4.35, SD = 2.19) were generally higher than statements presented in the strongly negative context and the weakly negative context (M = 3.08, SD = 1.56; M = 3.39, SD = 1.73, respectively). The main effect of Statement was also significant ($F_1(1, 88) = 238.92$, $p < .001$, MSE = 1.92; $F_2(1, 11) = 55.18$, $p < .001$, MSE = 1.12). The politeness ratings for the ironic statements (M = 4.56, SD = 1.96) were significantly higher than those for the literal statements (M = 2.71, SD = 1.39).

Confidence Ratings

For the confidence ratings, there was a significant Context x Statement interaction ($F_1(1, 88) = 14.31$, $p < .001$, $MSE = .48$; $F_2(1, 11) = 11.07$, $p < .01$, $MSE = .10$). The nature of this interaction was revealed with planned comparisons. The comparison of statements across the strongly negative context showed that participants were marginally more confident in their ratings for the literal statements than they were for the ironic statements ($t_1(88) = 1.87$, $p = .06$, $SE = .11$; $t_2(11) = 1.70$, $p > .05$, $SE = .10$). The comparison of statements across the weakly negative context showed that participants were equally confident in their ratings for the literal and ironic statements ($t_1(88) = 1.53$, $p > .05$, $SE = .11$; $t_2(11) = 1.76$, $p > .05$, $SE = .11$). The comparison of statements across the neutral context showed that participants were more confident in their ratings for ironic statements than they were for literal statements ($t_1(88) = 6.90$, $p < .001$, $SE = .12$; $t_2(11) = 6.00$, $p < .001$, $SE = .14$).

The main effect of Context was significant ($F_1(1, 88) = 3.81$, $p = .05$, $MSE = .48$; $F_2(1, 11) = 1.55$, $p > .05$, $MSE = .17$). Participants were generally more confident in their ratings for the statements presented in the strongly negative contexts ($M = 5.47$, $SD = 1.08$) than they were for the statements presented in the weakly negative contexts and the neutral contexts ($M = 5.27$, $SD = 1.18$; $M = 5.32$, $SD = 1.38$, respectively). The main effect for Statement was also significant ($F_1(1, 88) = 5.57$, $p < .05$, $MSE = .61$; $F_2(1, 11) = 5.24$, $p < .05$, $MSE = .09$). The confidence ratings for the ironic statements ($M = 5.46$, $SD = 1.21$) were higher than they were for the literal statements ($M = 5.25$, $SD = 1.24$), and as noted above, this difference was largely driven by the neutral condition.

Memory Scores

In scoring the free recall task, a score of 1 was given for each of the 12 items correctly recalled and a mean proportion recalled was calculated for each condition. A strict scoring system was used so that no errors in wording were allowed. The Context x Statement interaction was not significant ($F_1 < 1$; $F_2 < 1$). The comparisons of statements across all context conditions also showed no differences in recall of the statements by subjects or by items ($p > .05$). As well, the main effect of Context was not significant ($F_1 < 1$; $F_2 < 1$). The main effect of Statement, however, was significant ($F_1(1, 88) = 6.43$, $p = .01$, $MSE = .09$; $F_2(1, 11) = 2.89$, $p > .05$, $MSE = .04$), such that the memory scores for the literal statements ($M = .44$, $SD = .50$) were higher than those for the ironic statements ($M = .37$, $SD = .48$). If these memory scores can be taken as an indirect indicator of extent of processing, then they provide no evidence that ironic statements are more difficult to interpret.

This experiment provided valuable information about how the target statements were interpreted. In order to investigate processing for the target statements, reading times of the statements were examined in Experiment 3a.

Experiment 3a

Purpose

This experiment was designed to investigate the reading times for literal and ironic statements. Certain theories of irony (e.g., Gibbs, 1994) posit that reading times for literal and ironic statements should be equal, while other theories of irony (e.g., Giora, 1995) suggest that ironic statements will take longer to read than literal statements. The

purpose of this experiment was to determine whether the relative speed of processing for literal and ironic statements depends on strength of context. That is, a context that is strongly biasing towards irony (i.e., strongly negative) might lead to equivalent reading times for ironic and literal statements, while a context that is weakly biasing towards irony (i.e., weakly negative) might lead to longer reading times for ironic than for literal statements.

Method

Participants

There were 48 (27 female and 21 male) undergraduate students from the University of Calgary, aged 17 to 37 (M = 20.79, SD = 3.59), who participated in this study. Participation was voluntary and participants received a bonus credit towards a psychology course in exchange for their participation. As in Experiment 2, there were six versions of the stimuli. In this experiment, eight participants were presented with each version.

Materials

The stimuli for this experiment consisted of 10 practice trials, 12 experimental trials, and 63 filler trials. The filler trials were the same fillers used by Pexman et al. (2000). The 12 experimental trials were the same 12 stimuli sets as in Experiment 2, with a wrap-up sentence added after each target statement. The wrap-up sentence was included to catch any processing that occurred after the target statement had been read (i.e., the spillover effect). If the reader has not completely resolved the meaning of the target statement and it requires more time to process, this will lead to longer reading times in

the wrap-up sentence. Each wrap-up sentence was eight words long. For example, for the stimulus set with the target statement, Jason is a poor driver, the wrap-up sentence was April and Lynn talk while drinking their coffee.

To ensure that participants were attending to the stimuli they were reading, simple yes/no comprehension questions were asked after each paragraph was read. The questions did not involve assessment of speaker intent for the target statements.

The stimuli were presented on a 17-in Sony monitor controlled by a Macintosh G3 and presented using PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993) in a one-word-at-a-time moving window format. A PsyScope button box was used to advance the text and respond to the comprehension questions. The button box recorded the reading times for each word (the interval between successive button presses) with millisecond accuracy.

Procedure

Participants were tested individually. Each paragraph was presented on the computer screen, as a series of dashes for each word, with a space separating each word. The target statements were always presented on one line on the computer screen so that participants did not switch lines in the middle of the target statement.

Participants pressed the middle button on the button box to reveal the first word in the paragraph. Each subsequent button press revealed the next word, replacing the previous word with dashes again. The reading time for each word was recorded as well as processing time for the space between each sentence.

Each paragraph was read in this manner and a comprehension question was

answered after each paragraph by pressing the left or right button on the button box. Participants were instructed to read at their own normal reading pace. They were then presented with 10 practice trials to familiarize them with the reading style and were given time to ask any questions about the procedure.

Results and Discussion

The data in this experiment were reading times for each word in the target statements and the wrap-up sentences. There were 14 reading locations that were examined: the five words in each target statement, the space after the target statements, and the eight words in each wrap-up sentence (e.g., Jason is a poor driver. [space] April and Lynn talk while drinking their coffee). Mean reading times for all locations are presented in Figure 1. To test whether context condition affected the reading times of ironic and literal statements, 3 (context) x 2 (statement) repeated-measures ANOVAs were conducted for each location. We also conducted planned comparisons (which were simple main effects for the interaction) for statement effects within each context condition. Before analyses were conducted, 12.5% of the data were excluded: 1.5% for reading times shorter than 100 ms or longer than 2000 ms and 11% for wrong answers on the comprehension questions. Of the 14 reading locations, 4 locations showed significant effects (see Table 2 for mean reading times for these locations). Results for these four locations are presented separately.

Fifth Word in Target Statement

For this location, there was a significant Context x Statement interaction ($F_1(1, 47) = 5.79$, $p < .05$, $MSE = 27313.21$; $F_2(1, 11) = 2.34$, $p > .05$, $MSE = 14084.32$).

Planned comparisons for the ironic and literal statements, across the strongly negative context, showed that reading times for the ironic statements were marginally slower than those for the literal statements ($t_1(47) = 1.84$, $p = .07$, $SE = 32.58$; $t_2(11) = 1.32$, $p > .05$, $SE = 45.73$). The comparison of statements across the weakly negative context showed the opposite effect: reading times for the ironic statements were significantly faster than the literal statements ($t_1(47) = 2.07$, $p < .05$, $SE = 30.39$; $t_2(11) = 1.46$, $p > .05$, $SE = 39.58$). The comparison of statements across the neutral context showed equivalent reading times for ironic statements and literal statements, ($t_1 < 1$; $t_2 < 1$).

The main effect of Context was not significant ($F_1 < 1$; $F_2 < 1$). The main effect of Statement was also not significant ($F_1 < 1$; $F_2 < 1$).

Sixth Word in Wrap-up Sentence

For this location, the Context x Statement interaction was significant only by items, ($F_1(1, 47) = 1.66$, $p > .05$, $MSE = 7254.02$; $F_2(1, 11) = 5.89$, $p < .05$, $MSE = 1210.73$). The comparison of ironic and literal statements, across the strongly negative context, showed that following the ironic statements the reading times were slower at this location than they were following the literal statements ($t_1(47) = 2.46$, $p < .05$, $SE = 25.40$; $t_2(11) = 3.19$, $p < .01$, $SE = 21.46$). The comparison of statements across the weakly negative context showed equivalent reading times for this location following the ironic statements and the literal statements ($t_1 < 1$; $t_2 < 1$), as did the comparison for statements across the neutral context, ($t_1 < 1$; $t_2(11) = 1.48$, $p > .05$, $SE = 17.33$).

The main effect of Context was not significant ($F_1(1, 47) = 1.26$, $p > .05$, $MSE = 8074.01$; $F_2(1, 11) = 1.48$, $p > .05$, $MSE = 2393.14$). The main effect of Statement was

significant ($F_1(1, 47) = 6.96$, $p < .05$, MSE = 7856.40; $F_2(1, 11) = 12.99$, $p < .01$, MSE = 1501.96). The nature of this effect was that, in general, the reading times for this location following the ironic statements (M = 353.33 ms, SD = 158.53) were longer than those following the literal statements (M = 322.28 ms, SD = 123.95).

Seventh Word in Wrap-up Sentence

For this location, the Context x Statement interaction was not significant ($F_1 < 1$; $F_2(1, 11) = 2.68$, $p > .05$, MSE = 1109.61). The comparison of statements across the strongly negative context, however, showed that the reading times for this location following the ironic statements were longer than those following the literal statements ($t_1(47) = 2.03$, $p < .05$, SE = 24.68; $t_2(11) = 2.75$, $p < .05$, SE = 20.17). The comparison of statements across the weakly negative context showed equivalent reading times for this location following the ironic statements and the literal statements ($t_1 < 1$; $t_2 < 1$). The comparison of statements across the neutral context also showed equivalent reading times for this location ($t_1 < 1$; $t_2 < 1$).

The main effect of Context was not significant ($F_1(1, 47) = 2.66$, $p > .05$, MSE = 7526.44; $F_2(1, 11) = 3.47$, $p > .05$, MSE = 2151.48). The main effect of Statement was only significant by items ($F_1(1, 47) = 2.75$, $p > .05$, MSE = 11421.74; $F_2(1, 11) = 5.88$, $p < .05$, MSE = 1879.30). In general, the reading times for this location following the ironic statements (M = 353.95 ms, SD = 169.74) were longer than those following the literal statements (M = 328.20 ms, SD = 114.37).

Eighth Word in Wrap-up Sentence

For this location, the Context x Statement interaction was not significant ($F_1 < 1$;

$F_2 < 1$). The comparison of statements across the strongly negative context, however, showed that the reading times for this location following the ironic statements were longer than those following the literal statements ($t_1(47) = 2.20$, $p < .05$, SE = 38.07; $t_2(11) = 2.04$, $p = .07$, SE = 39.43). The comparison of statements across the weakly negative context showed equivalent reading times for this location following the ironic statements and the literal statements ($t_1 < 1$; $t_2 < 1$). The comparison of statements across the neutral context also showed equivalent reading times for this location ($t_1 < 1$; $t_2 < 1$).

The main effect of Context was significant ($F_1(1, 47) = 4.25$, $p < .05$, MSE = 46956.98; $F_2(1, 11) = 2.98$, $p > .05$, MSE = 17780.07). In general, the reading times for this location for the strongly negative context (M = 486.68 ms, SD = 292.06) and the weakly negative context (M = 453.62 ms, SD = 274.84) were faster than the neutral context (M = 539.57 ms, SD = 371.34). The main effect of Statement was not significant ($F_1 < 1$; $F_2 < 1$). This finding of additional processing required for an ironic statement after the target statement has been read replicates similar findings in past research (e.g., Pexman, et al., 2000; Pexman & Olineck, 2001).

Gibbs (1994) and Giora (1995) argue two different views of how irony is processed. Gibbs argued that an ironic statement can be processed as ironic without activating a literal interpretation first, and will, therefore, take no longer to read than a literal statement. Giora argued that an ironic statement cannot be processed as ironic without activating a literal interpretation first, and will, therefore, take longer to read than a literal statement.

The results of this experiment showed that processing of ironic statements

depends on the context it follows. Opposite to what was hypothesized, the ironic statements in the strongly negative context took longer to process than the literal statements in the same context. But in the weakly negative context, ironic statements and literal statements were processed equally as fast. These results do not provide any information as to what meaning (or meanings) is being activated. Experiment 4b was conducted to determine the meaning (literal or ironic) that was activated when the target statement was read.

While the results are in the opposite direction to what was predicted, they do show that, as predicted, context plays a role in how irony is processed. In some cases (weakly negative context), an ironic statement is processed in the same time as a literal statement, while in other cases (strongly negative context) it takes longer to process. One possible reason for the unexpected direction of results may be participants' expectations. The strongly negative contexts were very negative, and in those contexts participants may have expected a direct insult to be used (instead of irony). In a very negative situation, it could be socially acceptable to use a direct insult. This expectation may have interfered with the processing of the ironic statements in the strongly negative contexts, leading to longer reading times. This "expectations" hypothesis was tested in Experiment 3b.

Experiment 3b

Purpose

The purpose of this experiment was to determine if people expect literal (or figurative) language to be used in certain situations over figurative (or literal) language.

Method

Participants

There were 54 (34 female and 20 male) undergraduate students from the University of Calgary, aged 18 to 28 (M = 20.74, SD = 2.74), who participated in this study. Participation was voluntary and all participants received bonus credit towards a psychology course in exchange for their participation. There were three versions of the stimuli and 18 participants were presented with each version.

Materials

The stimuli were the same 12 sets of stimuli used in the previous experiments. These stimuli were presented here in a paper and pencil task. First, the context paragraphs were pre-recorded on an audio-tape. The narrator read each paragraph, then both statements in the pair, with either an ironic intonation or a literal intonation as appropriate.

Procedure

Participants were tested in groups of up to 15 individuals. They were presented with a booklet containing 12 items. One context type from each set was presented in the booklet, followed by both statements for that set. Contexts in the booklet were presented in random order. Participants were instructed to listen to the tape and to read along in the booklet. When the paragraph was finished and the statement choices were read, they were asked to place an X beside the statement that they would expect to follow the paragraph.

Results and Discussion

The mean number of expected ironic and literal statements were calculated for

each context condition. For the strongly negative context, literal statements (M = 2.52 (out of 4), SD = 1.14) were expected more often than ironic statements (M = 1.48, SD =1.14), $t(53) = 3.33$, $p < .01$, SE = .31. For the weakly negative context, literal and ironic statements were expected equally as often (M = 2.20, SD = 1.25; M = 1.80, SD = 1.25, respectively), $t(53) = 1.20$, $p > .05$, SE = .34. In the neutral context, the literal and ironic statements were also expected equally often (M = 2.00, SD = 1.36; M = 2.00, SD = 1.36, respectively), $p = 1.00$. These results support the "expectations" hypothesis: a literal criticism is more likely to be expected in the strongly negative context. This expectation may be what caused the longer reading times in this condition as it may have caused interference with the processing of the statement. As there is no expectation for a literal statement over an ironic statement for the weakly negative or neutral contexts, no such interference was observed.

Experiment 4a

Purpose

Experiment 4 was designed to be very similar to the experiment conducted by Giora et al. (1998). The aim, however, was also to correct some potential problems in their methodology. There were two problems evident: (1) there were no neutral probe words used; and (2) there were no baseline data obtained for the probe words used. To correct these problems, I included neutral probe words and designed Experiment 4a to investigate whether the probe words to be used in Experiment 4b produced equivalent response times in an LDT.

Method

Participants

There were 10 undergraduate and graduate students (6 female and 4 male) from the University of Calgary, aged 18 to 26 (M = 21.70, SD = 2.91) who participated in this experiment. Participation was voluntary and undergraduate students received bonus credit towards a psychology course in exchange for their participation.

Materials

The word stimuli for this experiment included three sets of words: (1) 12 positive words; (2) 12 negative words; and (3) 12 neutral words. The positive and negative words were chosen from Experiment 2; they related to either the ironic meaning of the statements or the literal meaning of the statements. They were the words that were written down by the most number of participants in Experiment 2. The neutral words were not related to the contexts or statements and were matched with the positive and negative words for word length and for frequency based on Kučera and Francis' (1967) norms. The nonword stimuli were 36 nonwords chosen from a set used by Gibbs and Van Orden (1998).

The stimuli were presented on a 17-in Sony monitor controlled by a Macintosh G3 and presented using PsyScope and a PsyScope button box was used to respond to the words.

Procedure

The participants were tested individually. The words were presented randomly to participants. They were presented with a fixation point in the middle of the screen and

then a letter string appeared. Participants were instructed to decide if the letter string was a proper English word by pressing the right button on the button box and if the letter string was not a word by pressing the left button on the button box. They were asked to make the decision as quickly and accurately as possible.

Results and Discussion

Before any analyses were conducted, 2.5% of the data were excluded because of an incorrect response. The mean RT for each word type was calculated and planned comparisons were used to determine if there were differences between the word types. Bonferroni adjustments were used to control for alpha inflation across these comparisons.

The results showed that the positive words (M = 505.25 ms, SD = 98.06) were responded to more quickly than the negative words (M = 585.78 ms, SD = 143.37), $t_1(9) = 4.44$, $p < .01$, $SE = 18.13$; $t_2(11) = 3.40$, $p < .01$, $SE = 23.16$, and more quickly than the neutral words (M = 539.33 ms, SD = 92.17), $t_1(9) = 3.27$, $p = .01$, $SE = 10.42$; $t_2(11) = 2.29$, $p < .05$, $SE = 21.32$. The neutral words were responded to marginally faster than the negative words, ($t_1(9) = 2.28$, $p < .05$, $SE = 20.35$; $t_2(11) = 1.17$, $p > .05$, SE = 25.66). Because there were differences in RTs between the word types, the mean RTs for each word from this experiment were used as a baseline to calculate effect sizes in Experiment 4b.

Experiment 4b

Purpose

This experiment was a priming experiment designed to determine the time course of meanings activated for ironic statements. The experiment addressed whether or not a

person reads an ironic statement with a literal interpretation first, whether literal and ironic meanings are processed in parallel, or whether an ironic interpretation is activated alone. Generally, it was expected that the RTs for the probe words in this experiment will be slower than in Experiment 4a. This is because, in Experiment 4b, a context and statement were read, and then participants were asked to do an LDT on the probe word. This is effectively task-switching, where reading the contexts is the primary task and doing the LDT is the secondary task. When it closely follows the primary task, the secondary task will likely take longer than if it was presented in isolation (a "carry over effect"). Thus, LDT responses in Experiment 4b were always expected to be slower than in Experiment 4a. The smallest slowing effect size will therefore indicate facilitation for the probe word. If the ironic statements are interpreted as ironic without activating a literal interpretation first, it is expected that probe words supporting the ironic meaning will have a smaller effect size than probe words supporting the literal meaning. If, however, ironic statements are interpreted as literal first, probe words supporting the literal meaning will have a smaller effect size than probe words supporting the ironic meaning.

There were two ISI conditions used: 100 ms and 1000 ms. An ISI of 100 ms was used to investigate which meaning is activated immediately after the statement is read. An ISI of 1000 ms was used to investigate meanings activated after additional processing.

Method

Participants

There were 81 (63 female and 18 male) undergraduate students from the

University of Calgary, aged 18 to 40 (M = 21.46, SD = 4.28), who participated in this study. Participation was voluntary and all participants received bonus credit towards a psychology course in exchange for their participation. There were 36 versions of the stimuli and two participants were presented with each version. Some versions received more than two participants in order to replace data that was excluded because of response errors or long RTs. There were seven participants excluded from analysis because over 50% of their data included incorrect responses or long RTs.

Materials

The same materials as in Experiment 3 were used in this experiment, with some changes. The paragraphs were again presented in a moving window format, however, instead of participants advancing each word, each word was presented on the screen for a fixed time of 250 ms. Therefore, the task was not self-paced as in Experiment 3.

The second change that was made was in the stimuli. The wrap-up sentence and comprehension question were removed from each paragraph. After the last word in the target statement, a probe word appeared either 100 ms or 1000 ms after the offset of the last word. The probe words were either literally related to the target statements or were ironically related to them. For the ironic statement Jason is a good driver, the ironically related probe word was dangerous and the literally related probe word was cautious. The literal statement for the same stimulus situation was Jason is a poor driver, and the literally related probe word was dangerous and the ironically related probe word was cautious. There were also neutral words (e.g., expanding) assigned to each context situation. Nonwords were also presented as probes for some of the filler trials, along with

related and unrelated probe words. Across all trials (targets and fillers) 50% of the probes were words and 50% were nonwords.

Procedure

Participants in this experiment were tested individually. They were presented with the paragraphs one word at a time. After the last word of the target statement was presented, the screen went blank for either 100 ms or 1000 ms and a letter string appeared in the centre of the screen. Participants were instructed to decide if the letter string was a word (by pressing the right button on the button box) or a nonword (by pressing the left button). They were asked to make the decision as quickly and accurately as possible.

Results and Discussion

Mean RTs were calculated. Before analyses were conducted, 1.5% of the data were excluded because of an incorrect response and 2% of the data were excluded because of response latencies over 1700 ms. Outliers in the data were also excluded (RTs larger than 3 standard deviations from the mean) which was another 2% of the data.

The mean effect size for each condition was calculated and these effect sizes were analyzed by items using separate 3 (Context) x 2 (Statement) x 3 (Probe Word Type) repeated measures ANOVAs for both ISI conditions. Planned comparisons were also conducted for probe word type effects within each Context x Statement condition. Bonferroni adjustments were used to control for alpha inflation across these comparisons. The data were not analyzed by subjects since the design was incomplete by subjects (there were more conditions than items). The consequence of an items analysis only is reduced power. Nonetheless, certain effects were observed. Recall that a small effect size is taken

to mean that the meaning for that word is currently activated, resulting in facilitation.

In the 100 ms ISI condition, the Context x Statement x Probe Word Type interaction approached significance, $F(1, 11) = 3.98$, $p = .07$, $MSE = 13862.53$, but in the 1000 ms ISI condition, the Context x Statement x Probe Word Type interaction was not significant, $F < 1$.

Strongly Negative Context

Ironic Statement. In the 100 ms ISI condition, the effect sizes for the ironic and literal probe words were statistically equivalent, $t(11) = 1.50$, $p > .05$, $SE = 48.62$ (see Figure 2). The effect size for the ironic probe words was marginally smaller than the effect size for the neutral probe words, $t(11) = 2.09$, $p = .06$, $SE = 74.05$. The effect sizes for the literal and neutral probe words were also equivalent, $t(11) = 1.26$, $p > .05$, $SE = 65.33$. These results suggest that at 100 ms, the ironic and literal meanings of the ironic statement in the strongly negative context condition are processed in parallel. Because the effect size for the ironic probe words was marginally smaller than the effect size for the neutral probe words, and the effect sizes for the literal and neutral probe words were equivalent, this indicates that the ironic meaning may be slightly more dominant than the literal meaning.

In the 1000 ms ISI condition, the effect size for the ironic probe words was equivalent to the effect sizes of the literal and neutral probe words ($t < 1$; $t(11) = 1.29$, $p > .05$, $SE = 40.54$, respectively). The effect sizes for the literal and neutral probe words were also equivalent, $t(11) = 1.12$, $p > .05$, $SE = 52.45$. Because all of the probe word types have equivalent effect sizes, this indicates that, by 1000 ms after the offset of the

last word in the target statement, the meaning of the statement was resolved and participants were equally able to respond to the words, across all probe word types.

Literal Statement. In the 100 ms ISI condition, the effect size for the literal probe words was equivalent to the effect size of the ironic and neutral probe words ($t < 1$; $t(11) = 1.19$, $p > .05$, SE = 42.38, respectively). The effect sizes for the ironic and neutral probe words were also equivalent, $t(11) = 1.47$, $p > .05$, SE = 34.24. The equivalence in effect sizes across probe word types indicates that, by this time, the meaning of the literal statement has already been resolved and there is no interference (or facilitation) between the meaning of the statement and the meaning of the probe words.

In the 1000 ms ISI condition, the effect size for the ironic probe words was marginally smaller than the effect size for the literal probe words and marginally smaller than the effect size for the neutral probe words ($t(11) = 2.21$, $p = .05$, SE = 42.80; $t(11) = 2.04$, $p = .07$, SE = 49.44, respectively). The effect sizes for the literal and neutral probe words were equivalent, $t < 1$. The smaller effect size for the ironic probe words is difficult to explain because priming for an ironic meaning of the statement was not expected. One possible way to explain these results is that, as in the 100 ms condition, processing of the meaning of the statement is completely finished. As there is a bias against negative words already present (slower processing for negative words), it is possible that the negative context with the negative statement and the negative (literal) probe word exacerbated that bias, causing longer RTs, and, therefore, larger effect sizes. Alternatively, one could view the small effect size for the positive (ironic) probe word as a release from the accumulated negativity from the context and statement. The social consequences of such

a negative situation paired with the negative statement (a direct insult) could lead participants to contemplate (mull over) the situation (e.g., I can't believe she said/did that). The positive probe words may lead to a release from the negativity and the thoughts involved with the social consequences, resulting in facilitation for the positive, ironic probe words. Obviously, this explanation is post hoc and would require further experimentation to be evaluated properly.

Weakly Negative Context

Ironic Statement. In the 100 ms ISI condition, the effect size for the ironic probe words was significantly smaller than the effect size for the literal and neutral probe words, ($t(11) = 5.56$, $p < .001$, $SE = 25.70$; $t(11) = 4.49$, $p = .001$, $SE = 55.55$, respectively). The effect sizes for the literal and neutral probe words were equivalent, $t(11) = 1.71$, $p > .05$, $SE = 60.90$ (see Figure 3). These results show that 100 ms after the offset of the final word in the target statement, the ironic meaning is activated. The results also provide evidence that the literal meaning is not activated, which is consistent with Gibbs' (1986; 1993; 1994) argument. If the literal probe words had a significantly smaller effect size than the neutral probe words, this would have provided evidence for activation of the literal meaning. As the effect sizes are the same, however, the results provided evidence that in some instances, the ironic meaning of a statement is processed without first activating the literal interpretation.

In the 1000 ms ISI condition, the results showed the same pattern, but the differences in effect sizes were smaller. The effect size for the ironic probe words was marginally smaller than the effect size for the literal probe words ($t(11) = 2.12$, $p = .06$,

SE = 40.54) and marginally smaller than the effect size for the neutral probe words ($t(11) = 2.52$, $p < .05$, SE = 30.46). The effect sizes of the literal and neutral probe words were equivalent, $t < 1$. These results indicate that the ironic meaning is still active at this point and being processed, while the literal meaning is still not active.

Literal Statement. The results in this condition showed similar effects as in the strongly negative context. In the 100 ms ISI condition, the effect size for the literal probe words was statistically equivalent to the effect sizes of the ironic and neutral probe words ($t(11) = 1.06$, $p > .05$, SE = 63.87; $t < 1$, respectively). The effect size for the ironic probe words was marginally smaller than the effect size for the neutral probe words, $t(11) = 1.96$, $p = .08$, SE = 45.28. These effects are similar to the effects found in the strongly negative context in the 1000 ms ISI condition. It is again suggested that processing of the meaning of the literal statement is complete at this point and there is no interference from the particular meaning of the statement. The literal probe word is believed to have a slightly larger effect size, again, because the word is negative and is placed after a negative context and negative statement. The effect is smaller in this case, compared to the strongly negative context, most likely because the context is not as negative.

In the 1000 ms ISI condition, the effect size for the literal probe words was equivalent to the effect sizes for the ironic and neutral probe words ($t(11) = 1.40$, $p > .05$, SE = 49.01; $t < 1$, respectively). The effect sizes for the ironic and neutral probe words were also equivalent, $t(11) = 1.23$, $p > .05$, SE = 42.98. These results, again, support the idea that the processing of the literal statements was complete at this time and thus there was no interference between the meaning of the statements and the probe words.

Neutral Context

Ironic Statement. In the 100 ms ISI condition, the effect size for the literal probe words was marginally smaller than the effect size for the ironic probe words, $t(11) = 1.77$, $p = .10$, $SE = 59.10$ (see Figure 4). The effect sizes for the literal and ironic probe words were equivalent to the effect size for the neutral probe words ($t < 1$; $t(11) = 1.50$, $p > .05$, $SE = 70.85$, respectively). In this condition, the ironic statement was interpreted in a literal manner, as indicated by the ratings in Experiment 2. Therefore, the literal meaning (probe word) should be activated. These results are similar to those for the literal statements in the strongly negative and weakly negative contexts. The effect size of the literal probe words was marginally smaller than the effect size of the ironic probe words, but is equivalent to the effect size of the neutral probe words. This may indicate that processing of the statements was almost complete, but the literal meaning was still slightly active, causing some interference for the ironic and neutral probe words.

In the 1000 ms ISI condition, the effect size for the ironic probe words was equivalent to the effect sizes of the literal and neutral probe words ($t < 1$). The effect sizes for the literal and neutral probe words were also equivalent, $t(11) = 1.40$, $p > .05$, $SE = 63.36$. The results in this ISI condition provide evidence that, at least in this context condition, processing of the statements is complete at this time since the effect sizes for all the probe word types were equivalent.

Literal Statement. In the 100 ms ISI condition, the effect size for the ironic probe words was equivalent to the effect sizes for the literal and neutral probe words ($t(11) = 1.35$, $p > .05$, $SE = 45.54$; $t(11) = 1.35$, $p > .05$, $SE = 45.73$, respectively). The effect sizes

for the literal and neutral probe words were also equivalent, $t < 1$. Although the ratings data from Experiment 2 indicated that this statement type was interpreted to be more sarcastic than the "ironic" statements, the mean sarcasm rating was low ($M = 2.84$, $SD = 1.51$) indicating that these statements are only occasionally being interpreted as ironic statements. This is consistent with the equivalence in the effect size across all three probe word types. The results in this condition are similar to the literal statements in the strongly negative and weakly negative contexts, indicating that no particular meaning of the statement is dominant.

In the 1000 ms ISI condition, the effect size for the ironic probe words was equivalent to the effect sizes of the literal and neutral probe words ($t(11) = 1.05$, $p > .05$, $SE = 55.20$; $t < 1$, respectively). The effect sizes for the literal and neutral probe words were equivalent, $t < 1$. Again, this provides evidence that no meaning is dominant for these potentially ironic compliments.

In summary, the results of this experiment provide evidence that when processing ironic statements, a literal meaning is not activated first, as Giora et al. (1998) argued. Instead, the results supported Gibbs' (1986; 1993; 1994) view of ironic processing. The results showed that in some contexts (the strongly negative in this experiment), the literal and ironic meanings of an ironic statement are processed in parallel, while in other contexts (the weakly negative), the ironic meaning of a statement is processed alone.

The priming results also showed that ironic and literal statements are processed differently. When we interpret literal language, we are able to resolve the meaning immediately after the statement is read/heard. For ironic language, however, the meaning

is not resolved immediately after it is read/heard, but shortly thereafter.

General Discussion

Gibbs (1986; 1993; 1994) and Giora (1995; Giora et al., 1998) argue two different views of ironic processing. Examining reading times, Gibbs (1986) found that ironic statements were processed faster than nonironic statements. Giora at al., however, found that ironic statements take more time to process than literal statements. In the current research, I found that the preceding context affected how an ironic statement was processed. When the preceding context was strongly negative, the ironic statements took longer to process than the literal statements (apparently because there was activation of both literal and ironic meanings), which supported the results found by Giora et al. When the preceding context was weakly negative, however, the ironic statements were processed somewhat more quickly than the literal statements (apparently because only ironic meanings were activated), which supported Gibbs' findings.

Traditionally, reading time studies have involved presenting an entire statement to participants and measuring the time involved in reading (or making a judgement about) the entire statement. The current study investigated processing for each word participants read in a moving window paradigm. This improved upon previous research because, by examining processing for each word read, it is possible to determine where any extra processing time occurred. For example, if the extra time was taken on the first two words of the statements, but the endings of the statements were processed in the same time, this difference would not be due to the processing of literal versus ironic statements as the reader could not yet know if the statement is literal or ironic.

Unfortunately, reading times alone do not provide enough evidence to prove that an ironic statement is processed as ironic or whether a literal interpretation is processed first or in parallel. Reading times indicate only that a particular type of statement required more processing time. The results of Experiment 3b indicated that people expect a literal criticism to follow a strongly negative context. Is the expectation which is not met the reason for longer processing times?

The results from Experiment 4 supported this possibility. For the ironic statements in the strongly negative contexts, results indicated that the literal meaning of an ironic statement is not processed first as there was equivalent facilitation for the ironic and literal probe words (vs neutral), even at the short (100 ms) ISI. The results provided evidence supporting Gibbs' (1986; 1993; 1994) view, that the ironic and literal meanings of the statement are processed in parallel, with the ironic meaning showing some dominance over the literal meaning.

The results from the weakly negative context, however, provide evidence that, in this context type, ironic statements are processed as ironic, <u>without</u> any evidence of processing for the literal meaning. In this context, the effect size of the ironic probe words was smaller than that for the literal and neutral probe words, with the effect sizes for the literal and neutral probe words being equivalent. The context facilitated the ironic meaning of the statement and not the literal meaning. This is in opposition to what Giora et al. (1998) found. In their study, the RTs for the ironic probe words were always longer than the RTs for the literal probe words. It was not until 2000 ms after the offset of the final word in the target statement that they found equivalent RTs for the ironic and literal

probe words. Yet, at an ISI of 100 ms, the results of the present study showed facilitation for the ironic probe words.

Two problems became evident with the results found by Giora et al. (1998), making a direct comparison of the studies impossible. First, in this study, the results of Experiment 4a showed that the probe words were not equivalent to each other. In terms of LDT responses, the positive words, which were the literal probe words for the ironic statements, were responded to more quickly than the negative (ironic) probe words. Thus, it was necessary to calculate effect sizes in Experiment 4b using the baseline responses from Experiment 4a. This allowed for a more accurate comparison to be made between the words, examining the effects of the context and statements alone. Giora et al. only examined the RTs for the probe words and did not obtain any baseline data. The words were similar to those used in the present study, in that the ironic probe words were negative and the literal probe words were positive. Thus, their finding of faster RTs for the literal probe words could be attributed to this baseline difference.

A second problem was that Giora et al. (1998) did not use neutral probe words. Without a comparison to make against a neutral word which does not relate to the text, it is difficult to determine exactly what meanings are activated from the statement. For example, if response times for the literal and ironic probe words do not differ, it cannot be determined whether both meanings are activated, or if neither is activated. However, if the ironic and literal probe words are both responded to faster than the neutral probe words, this indicates that both the ironic and literal meanings are activated.

There are some limitations to Experiment 4b and the conclusions that can be

drawn from it. First of all, the design was incomplete. There were more conditions than there were test items, which prevented an analysis by subjects. While the items analysis showed some significant effects, there may not have been enough power to show other differences that may have occurred. Secondly, the situation was quite complex. The LDT following the reading task measured incidental effects of processing the paragraphs that were read. Thus, the effects found for LDT responses may be a result of more than the processing of the paragraph just read. Finally, in much of the priming literature, the strength of the association between the prime and the target is measured in some way. In this experiment, however, the strength of the relationships between the target statements and the probe words was not measured. Thus, these relationships may vary across the target sentences, compromising the power of the analyses to detect significant effects.

The research presented here clearly showed that context affects that way irony is processed. In the strongly negative contexts, the reading times showed that the ironic statements took longer to process than the literal statements. When the priming results were examined, they indicated that, in that context, the ironic and literal meanings of the statements were processed in parallel. This, along with the failed expectations in the strongly negative context, could contribute to the longer reading times for ironic statements. When the statements followed the weakly negative contexts, the reading times showed that ironic statements were processed more quickly than the literal statements. The priming results indicated that, in that context, the ironic meanings of the statements were processed without the literal meaning being processed. This could explain why ironic statements took no longer to process than literal statements. These context effects

are well characterized by David Rumelhart (1993):

> Linguistic utterances are always interpreted in some context. The context of utterance, along with any knowledge available to the listener, may potentially be employed in the process of constructing an interpretation of the utterance. Moreover, I suspect that this knowledge is not employed in any ad hoc way, say simply as a filter in choosing among the various possible readings a sentence might have. Rather, these elements play a central role in determining what interpretations are possible for a given utterance (p. 76).

How is Irony Processed?

The priming data in Experiment 4b showed that literal statements are interpreted/resolved very rapidly. While the ironic statements can be processed as ironic without entertaining a literal meaning (in the weakly negative context), the meaning does not appear to be resolved as early as it is for a literal statement. The results indicate that a literal statement is processed differently from an ironic statement. The extra time required to comprehend an ironic statement may be a result of the ambiguity created by such a statement. While there is evidence supporting the idea that an ironic meaning is activated quite rapidly, the extra comprehension time may come in trying to fully resolve the subtleties of speaker intent. Kreuz et al. (1991) reported that irony fulfilled more communication goals (e.g., to mock, to be funny, to criticize) than literal language. Judging these aspects of speaker intent may take time.

In Experiment 3a, results for the weakly negative context showed that the literal and ironic statements had equivalent processing times. In Experiment 4b, however, the

priming data showed that the nature of processing for these two types of statements was quite different: rapid resolution of meaning for literal statements, while activation of meaning for ironic statements seemed to last longer. Thus, the processing for literal statements seems to involve something additional to the meaning resolution. It may be that the literal statements require additional processing (in addition to meaning resolution) because of the social consequences of the negative statements (violating the conversational norm of "if you don't have something nice to say, don't say anything at all"). The situation is somewhat different in the strongly negative context because in that context, readers expect a literal insult, not an ironic statement (see results of Experiment 3b), resulting in longer reading times for the ironic statement. The memory data in Experiment 2 may also be a result of this negativity. Memory tended to be better for the negative (literal) statements. There has been evidence that negative statements are better remembered than positive statements (e.g., Katz & Pexman, 1997; Kreuz et al., 1991). Thus, although meaning resolution seemed fast for literal statements, it may be that the negative nature of those literal statements (because of the strong social consequences of saying something so insulting) lead to additional processing time.

The priming data showed that the ironic meaning of an ironic statement takes longer to resolve than the literal meaning of a literal statement although overall processing time was equivalent (Experiment 3a). I believe that the ironic meaning takes longer because it requires more extensive integration (after the target has been read) to fully establish the speaker's intent. Pexman et al. (2000) have shown that the space after a target metaphor had a longer reading time when a metaphor was used ironically (as

compared to a situation where the metaphor was not used ironically). Also, Pexman et al. found that contextual cues (e.g., the occupation of the speaker) were integrated at the space after the target metaphor. There was no evidence that these cues were used when the targets were simply interpreted as metaphors. Thus, processing irony seems to involve immediate recognition of ambiguity (perhaps because of incongruity between the positive tone of words in the statement and the negative context), followed by extensive integration of all relevant cues in order to make a judgement about speaker intent. This is not the same as processing the literal meaning first, negating it, and switching to an ironic interpretation. The incongruity between statement and context could be marked by a single word in the statement (e.g., the modifier graceful in the target Tracy is a graceful dancer). If the context is negative (but not too strongly negative), then this incongruity will be detected even more rapidly, helping the comprehender to build a representation of the utterance that involves ironic intent on the part of the speaker.

These results support Gibbs' (1986; 1993; 1994) view of ironic processing to an extent, but also lend some support for the view held by Giora et al. (1998; Giora, 1995). These results also show that neither view is entirely correct. Gibbs (1986) argued that an ironic statement does not take longer to process than a literal statement and the ironic meaning can be processed alone, or in parallel with, the literal meaning of the ironic statement. Giora et al. argued that the literal meaning is always activated (unless the irony is highly familiar) before the ironic meaning. The results of the priming experiment showed that the ironic statement was processed without the literal meaning in the weakly negative context, and was processed in parallel with the literal meaning in the strongly

negative context. These results support Gibbs' view, but not Giora's view. The results from the reading time experiment lend support to both views of ironic processing. In the weakly negative context, the reading times for the ironic statements were equivalent to the reading times for the literal statements, supporting Gibbs' view of ironic processing. However, in the strongly negative context, the reading times for the ironic statements were longer than the reading times for the literal statements, supporting the view of Giora et al.

While the data from the current study support the processing time predictions of both Giora et al. (1998; Giora, 1995) and of Gibbs (1986; 1993;1994), I do not believe that it supports their ideas about why the increased or equivalent times occur. I believe that reading time data is influenced by a number of things, such as participants' expectations, and social consequences for the negativity of the situation, and not only by the processing of the ironic meaning of a statement or the processing of the literal meaning and then the ironic meaning. My results suggest that, in terms of meaning activation, an ironic statement can, if the context is supportive, be processed by direct activation of the ironic meaning. If the context is not supportive (e.g., if it is extremely negative, creating an expectation of a direct insult) then both literal and ironic meanings will be activated. Again, there were limitations in the experiments conducted and the conclusions that are drawn are not strong. Further experimentation needs to be done to strengthen the results to ensure that the conclusions and results are representative of the processing involved in irony. While the research was conducted in a lab setting and only used context cues, there are other cues which are generally present. With the presence of

other cues, I do not believe that the processing involved in irony would be any different, however, I think that the irony would be detected more easily and context itself may not show any effect on irony processing because of the ease in which irony would be detected.

The results of the present research suggest that processing of literal and ironic meanings is best explained by a model of comprehension that involves expectations derived from context, and a dynamic comprehension system that incorporates all available information in order to resolve ambiguity. Full elaboration of such a model will require additional experimentation.

References

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. Behavior Research Methods, Instruments, & Computers, 25, 257-271.

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. (1995). Muting the meaning: A social function of irony. Metaphor and Symbolic Activity, 10, 3-19.

Gerrig, R. J., & Goldvarg, Y. (2000). Additive effects in the perception of sarcasm: Situational disparity and echoic mention. Metaphor and Symbol, 15, 197-208.

Gerrig, R. J., & Healy, A. F. (1983). Dual processes in metaphor understanding: Comprehension and appreciation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 667-675.

Gibbs, P., & Van Orden, G. C. (1998). Pathway selection's utility for control of word recognition. Journal of Experiment Psychology: Human Perception and Performance, 24, 1162-1187.

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

Gibbs, R. W., Jr. (1987). Memory for requests in conversation revisited. American Journal of Psychology, 100, 179-191.

Gibbs, R. W., Jr. (1993). Process and products in making sense of tropes. In A. Ortony (Ed.). Metaphor and thought (2nd edition) (pp. 252-274). New York: Cambridge University Press.

Gibbs, R. W., Jr. (1994). Figurative thought and figurative language. In M. A. Gernsbacher (Ed.), Handbook of psycholinguistics (pp. 411- 446). New York: Academic Press.

Giora, R. (1995). On irony and negation. Discourse Processes, 19, 239-264.

Giora, R., & Fein, O. (1999). Irony: Context and salience. Metaphor and Symbol, 14, 241-257.

Giora, R., Fein, O., & Schwartz, T. (1998). Irony: Graded salience and indirect negation. Metaphor and Symbol, 13, 83-101.

Katz, A. N., & Pexman, P. M. (1997). Interpreting figurative statements: Speaker occupation can change metaphor to irony. Metaphor and Symbol, 12, 19-41.

Keil, F. C. (1986). Conceptual domains and the acquisition of metaphor. Cognitive Development,1, 73-96.

Kreuz, R., & 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., Long, D. J., & Church, M. B. (1991). On being ironic: Pragmatic and mnemonic implications. Metaphor and Symbolic Activity, 6, 149-162.

Kučera, H., & Francis, W. (1967). Computational analysis of present-day American English. Providence, RI: Brown University Press.

McElree, B., & Nordlie, J. (1999). Literal and figurative interpretations are

computed in equal time. Psychonomic Bulletin and Review, 6, 486-494.

O'Brien, E. J., & Myers, J. L. (1985). When comprehension difficulty improves memory for text. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 12-21.

Ortony, A., Schallert, D. L., Reynolds, R. E., & Antos, S. J. (1978). Interpreting metaphors and idioms: Some effects of context on comprehension. Journal of Verbal Learning and Verbal Behavior, 17, 465-477.

Pexman, P. M., Ferretti, T. R., & Katz, A. N. (2000). Discourse factors that influence on-line reading of metaphor and irony. Discourse Processes, 29, 201-222.

Pexman, P. M., & Olineck, K. M. (2001, July). Understanding positive and negative metaphor and irony: Off-line interpretation and on-line processing. Poster to be presented at the meeting of the Society for Text and Discourse, Santa Barbara, CA.

Pexman, P. M., & Olineck, K. M. (in press). Does sarcasm always sting? Investigating the impact of ironic insults and ironic compliments. Discourse Processes.

Rumelhart, D. E. (1993). Some problems with the notion of literal meanings. In A. Ortony (Ed.). Metaphor and thought (2nd edition) (pp. 71-82). New York: Cambridge University Press.

Shinjo, M., & Myers, J. L. (1987). The role of context in metaphor comprehension. Journal of Memory and Language, 26, 226-241.

Toplak M., & Katz, A. N. (2000). On the uses of sarcastic irony. Journal of Pragmatics, 32, 1467-1488.

Winner, E., & Gardner, H. (1993) Metaphor and irony: Two levels of

understanding. In A. Ortony (Ed.). Metaphor and thought (2nd edition) (pp. 425-443). New York: Cambridge University Press.

Winner, E., Windmueller, G., Rosenblatt, E., Bosco, L., Best, E., & Gardner, H. (1987). Making sense of literal and nonliteral falsehood. Metaphor and Symbolic Activity, 2, 13-32.

Appendix

Example Context and Statement Sets

Strongly negative, weakly negative, and neutral contexts appear, respectively, in italics.

1) Sam agreed to pick Christopher up after school. Sam never arrived to pick up Christopher and never called to say why / Sam arrived 1 hour late and apologized / Sam and Christopher talked about the dance on Friday. The next day Christopher is explaining to Jodi what happened. Christopher says:

Ironic statement: Sam is a nice friend.

Ironic probe word: irresponsible

Literal probe word: dependable

Literal statement: Sam is a rotten friend.

Literal probe word: irresponsible

Ironic probe word: dependable

Neutral probe word: dissolved

Wrap-up sentence: Christopher and Jodi were walking home from school.

2) Terri agreed to help Joan with the toy drive on Saturday. Terri spent the entire day socializing / Terri spent half the day socializing / Terri spent some of her time helping. The next day Joan is explaining to Sara what happened. Joan says:

Ironic statement: Terri is a super helper.

Ironic probe word: useless

Literal probe word: useful

Literal statement: Terri is a lazy helper.

Literal probe word: useless

Ironic probe word: useful

Neutral probe word: temple

Wrap-up sentence: The toy drive lasted from dawn until dusk.

3) Meredith told Kevin that she could memorize a 20 line poem in 5 minutes. Meredith started to recite the poem but forgot everything after the first 2 lines / Meredith started to recite the poem but only recited half before forgetting the rest / Meredith recited the poem and only forgot 2 lines. The next day Kevin is talking to Harry about the Poem. Kevin says:

Ironic statement: Meredith has a powerful memory.

Ironic probe word: forgetful

Literal probe word: amazing

Literal statement: Meredith has a weak memory.

Literal probe word: forgetful

Ironic probe word: amazing

Neutral probe word: disappear

Wrap-up sentence: Kevin and Harry were watching a funny movie.

	Strongly Negative Context		Weakly Negative Context		Neutral Context	
	Ironic Statement	Literal Statement	Ironic Statement	Literal Statement	Ironic Statement	Literal Statement
Ratings						
Sarcasm	5.74 (1.21)	1.87 (0.98)	4.85 (1.48)	2.18 (1.20)	1.85 (1.02)	2.84 (1.51)
Mocking	5.17 (1.20)	3.64 (1.64)	4.45 (1.49)	3.97 (1.30)	1.78 (0.97)	4.18 (1.50)
Politeness	3.35 (1.37)	2.82 (1.26)	4.13 (1.45)	2.71 (1.01)	6.14 (0.83)	2.52 (1.03)
Confidence	5.37 (0.92)	5.57 (0.87)	5.19 (0.96)	5.35 (0.90)	5.74 (0.92)	4.90 (1.14)
Free Recall	.33 (.31)	.42 (.38)	.39 (.38)	.43 (.35)	.38 (.37)	.46 (.39)

Table 1.

Mean Ratings and Free Recall Proportions (Standard Deviations in Parentheses) by Condition for Experiment 2

Table 2.

Correlations Between Ratings Dimensions and Free Recall Data for Experiment 2

Measure	1	2	3	4	5
1. Sarcasm ratings	----	.53**	-.06	-.01	-.11**
2. Mocking ratings	-.47**	----	-.36**	.02	-.01
3. Politeness ratings	-.31**	-.68**	----	-.17**	-.03
4. Confidence ratings	-.33**	-.31**	.29**	----	-.02
5. Free recall data	-.14**	.01	-.09	-.05	----

Note: Correlations above the diagonal are for the strongly negative and weakly negative contexts (df = 697) and correlations below the diagonal are for the neutral context (df = 355)

** $p < .01$

	Strongly Negative Context		Weakly Negative Context		Neutral Context	
	Ironic Statement	Literal Statement	Ironic Statement	Literal Statement	Ironic Statement	Literal Statement
Fifth word in	425.89	365.99	346.35	409.17	410.91	397.70
statement	(313.20)	(194.11)	(183.59)	(247.35)	(311.36)	(205.74)
Wrap-up 6	376.37	313.80	335.65	326.40	347.55	336.70
	(165.99)	(139.44)	(103.24)	(104.63)	(109.06)	(92.27)
Wrap-up 7	382.42	332.38	334.63	331.45	348.78	339.30
	(150.54)	(121.34)	(106.85)	(120.70)	(150.03)	(87.53)
Wrap-up 8	533.12	449.37	473.38	460.84	540.50	568.86
	(281.58)	(167.37)	(309.96)	(229.26)	(331.92)	(367.17)

Table 3.

Mean Reading Times (Standard Deviations in Parentheses) for Locations With Significance by Condition for Experiment 3a

Figure Captions

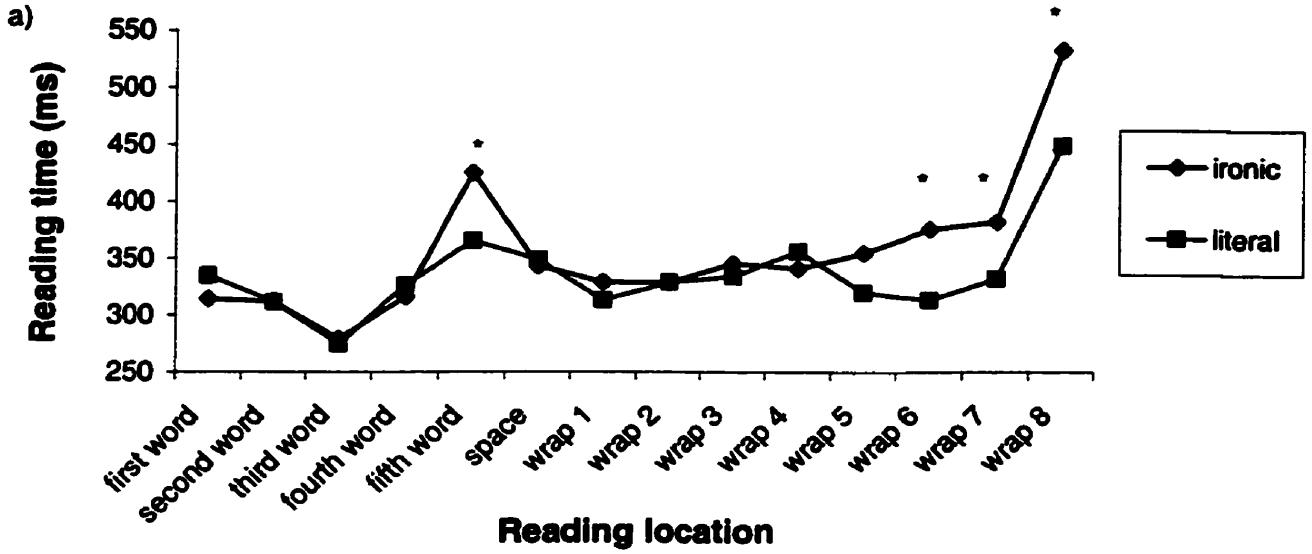
Figure 1. Mean reading times (in ms) for all reading locations for the (a) strongly negative context, (b) weakly negative context, and (c) neutral context.

Figure 2. Mean effect size for the ironic, literal, and neutral probe words across statement type and ISI condition, for the strongly negative context.

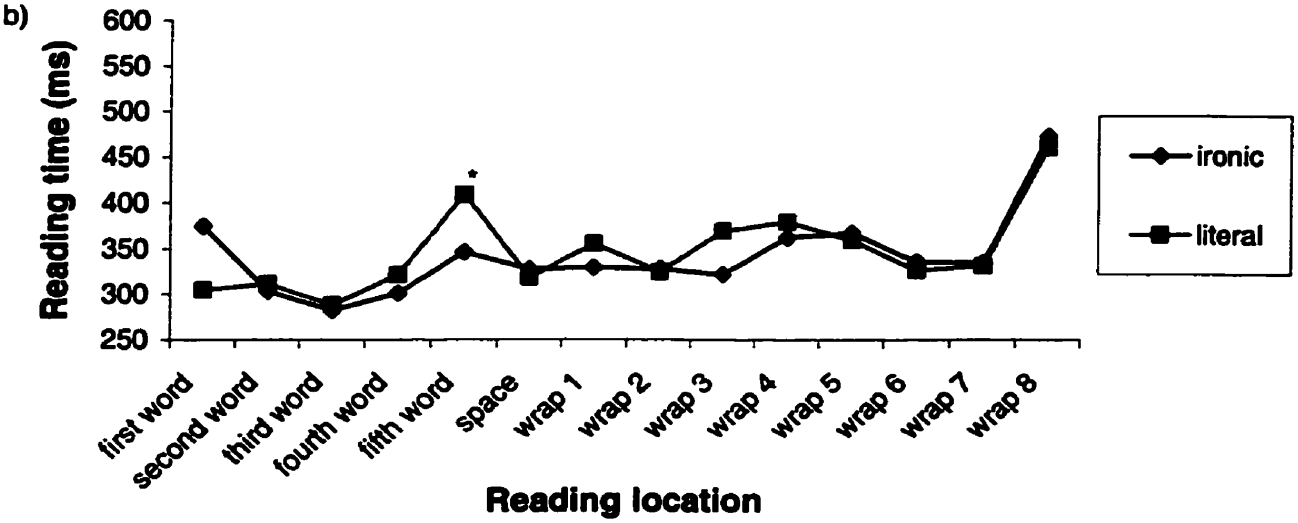
Figure 3. Mean effect size for the ironic, literal, and neutral probe words across statement type and ISI condition, for the weakly negative context.

Figure 4. Mean effect size for the ironic, literal, and neutral probe words across statement type and ISI condition, for the neutral context.

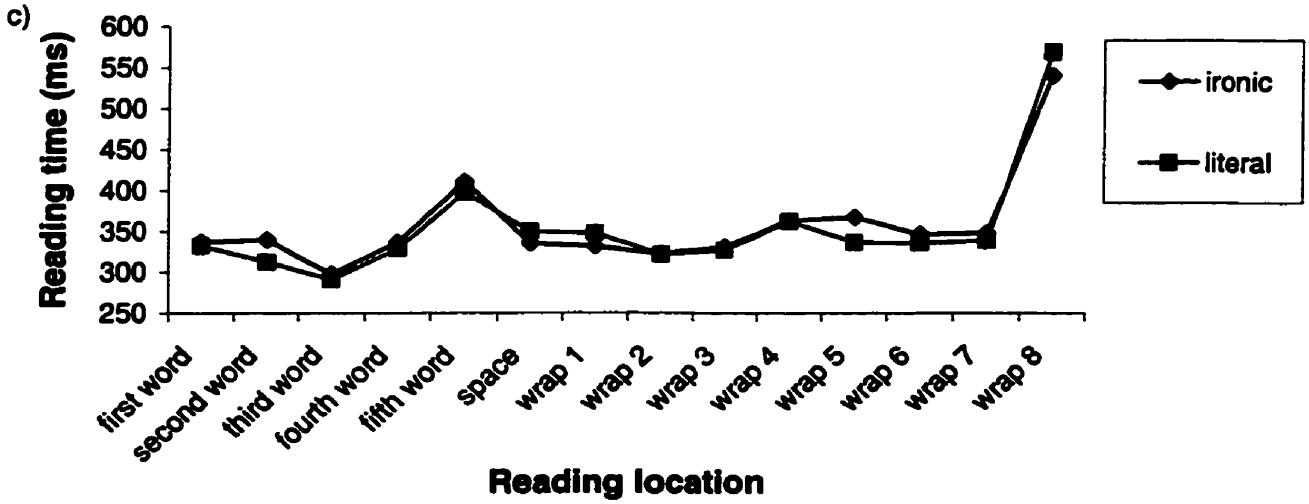
a)

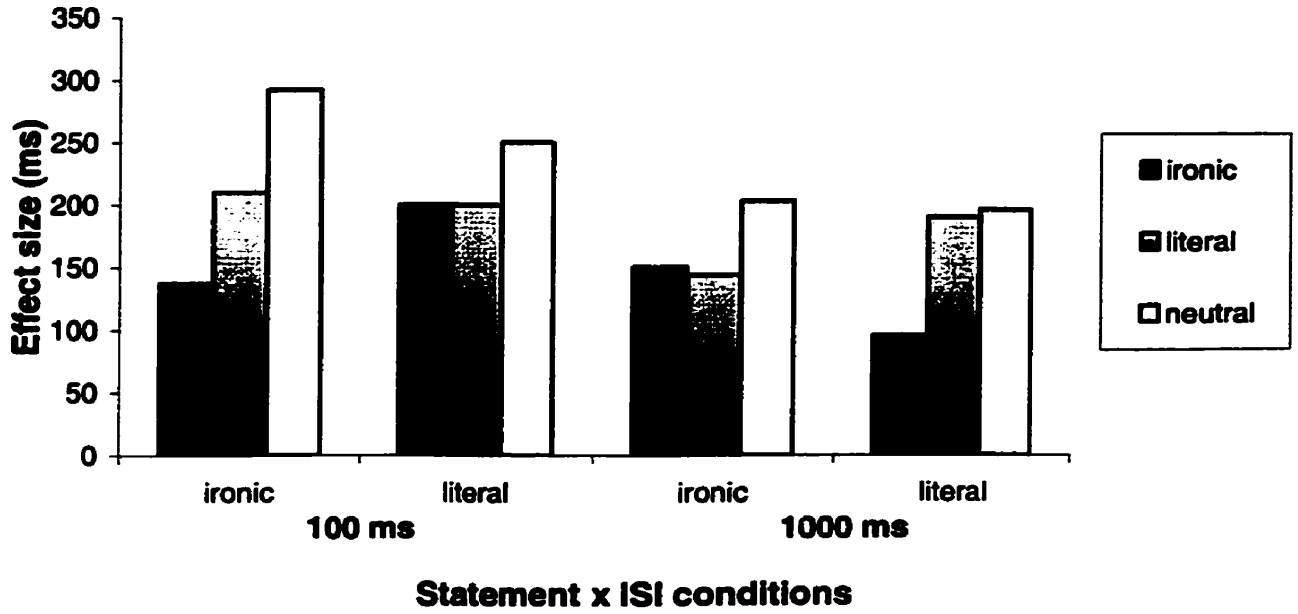

Reading time (ms)
550
500
450
400
350
300
250
ironic
literal
first word
second word
third word
fourth word
fifth word
space
wrap 1
wrap 2
wrap 3
wrap 4
wrap 5
wrap 6
wrap 7
wrap 8
Reading location

b)

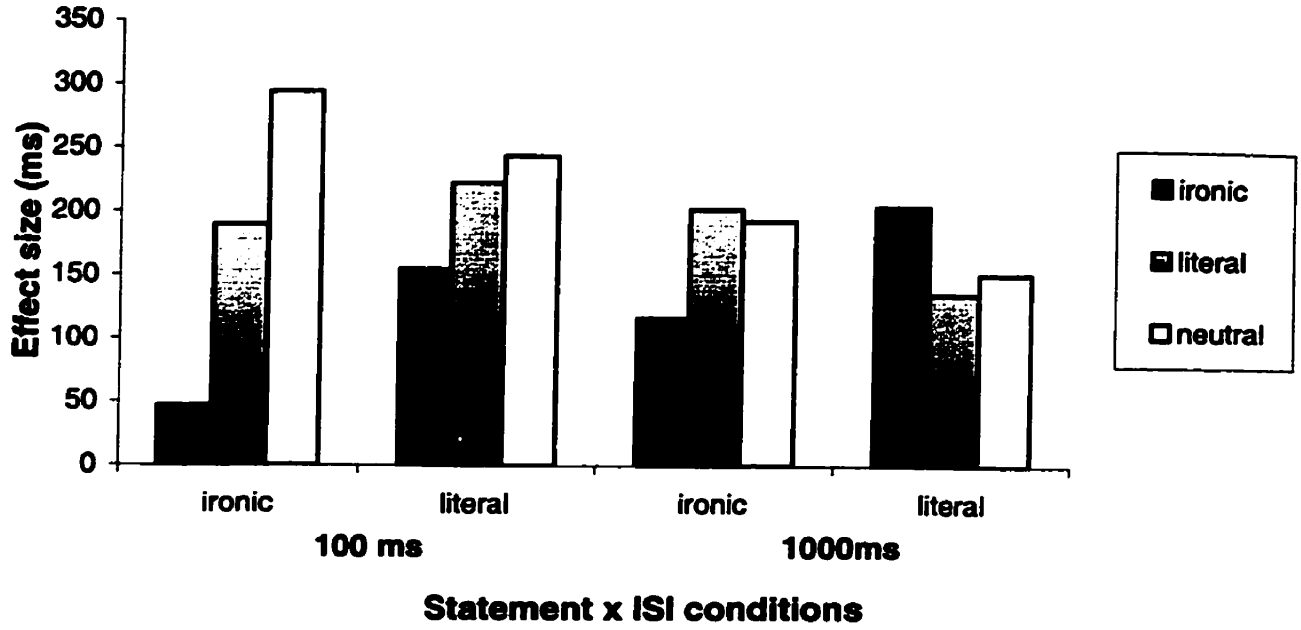

Reading time (ms)
600
550
500
450
400
350
300
250
ironic
literal
first word
second word
third word
fourth word
fifth word
space
wrap 1
wrap 2
wrap 3
wrap 4
wrap 5
wrap 6
wrap 7
wrap 8
Reading location

c)

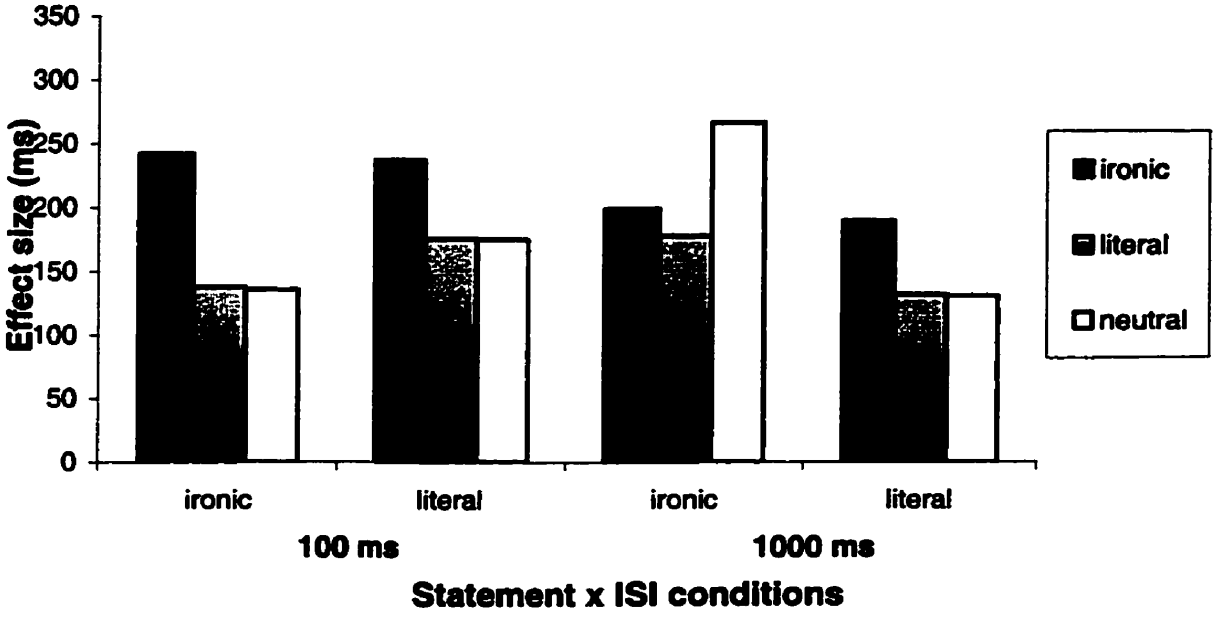

Reading time (ms)
600
550
500
450
400
350
300
250
ironic
literal
first word
second word
third word
fourth word
fifth word
space
wrap 1
wrap 2
wrap 3
wrap 4
wrap 5
wrap 6
wrap 7
wrap 8
Reading location



Statement x ISI conditions



Statement x ISI conditions


350
300
250
200
150
100
50
0
Effect size (ms)
ironic
literal
neutral
ironic
literal
ironic
literal
100 ms
1000 ms
Statement x ISI conditions