Ofer Fein*, Menahem Yeari and Rachel Giora

On the priority of salience-based interpretations: The case of sarcastic irony[1]

Abstract: Results from 2 experiments support the view that, regardless of strength of contextual expectation for utterance nonsalient (ironic) interpretation, (a) salience-based interpretations will not be blocked. Instead, they will be facilitated initially. And, (b) if conducive to the interpretation process, they will not be suppressed, albeit incompatible (Giora 2003; Giora and Fein 1999a; Giora and Fein 1999b; Giora et al. 2007). In Experiment 1, expectancy for an ironic utterance was manipulated by introducing an ironic speaker, whose ironic utterances were prefaced by overt ironic cues, making explicit the speaker's ironic intent. In Experiment 2, expectancy strongly biased via repeated exposure to ironic utterances, was further strengthened by informing participants that the experiment was testing sarcasm interpretation. Long processing times were allowed so as to tap later (suppression) processes. Results from reading times and lexical decisions support the temporal priority of salience-based interpretations, while arguing against both, the contextualist views (Gibbs 2002; Katz 2009) and the Gricean suppression hypothesis (Grice 1975).

Keywords: irony, sarcasm, salience, context, expectation hypothesis

1 Introduction

There is an enduring debate within psycholinguistics over the immediate effect of various sources of information on initial comprehension processes. An unresolved issue is whether it is contextual information that reigns supreme or whether it is coded information that enjoys priority. In this study we attempt to provide evidence that might tip the balance in favor of one of the views.

---

[1] Sarcasm and irony are used here interchangeably.

*Corresponding author: Ofer Fein:** Academic College of Tel-Aviv-Yaffo. E-mail: oferf@mta.ac.il
Menahem Yeari: Bar Ilan University, Israel.
Rachel Giora: Tel Aviv University.

1.1 Processing models: Activation, suppression, retention

According to context-based models, a strong context, supportive of the appropriate meaning, can govern comprehension processes very early on and facilitate contextually compatible meanings and interpretations immediately and exclusively. Such comprehension processes must therefore be *initially* frictionless, involving no contextually inappropriate meanings and interpretations, and hence no suppression of such interpretations (*the connectionist model*, e.g., Bates 1999; Bates and MacWhinney 1989; MacWhinney 1987; Small et al. 1988; *the direct access view*, *e.g.*, Gibbs 1986, Gibbs 1995, Gibbs 2002; *the constraint-based model*, e.g., Campbell and Katz 2012;[2] Katz 2009; McRae et al. 1998; Pexman et al. 2000). Accordingly, in the field of nonliteral language, context-based approaches, such as *the direct access view* (Gibbs 1979, Gibbs 1986, Gibbs 1994, Gibbs 2002), predict that strong contextual information will facilitate noncoded nonliteral interpretations immediately and exclusively; no recourse to suppression of irrelevant interpretations is anticipated (see also Keysar 1994; Ortony et al. 1978).

At the other end of the spectrum are lexicon-based approaches which assume an automatic response to a stimulus. Such response is sensitive only to relevant (e.g., linguistic) information. Initially, then, (adult) lexical access is impervious to context effects, which, however, often run parallel. According to one approach, *the modular view*, lexical access is exhaustive, making available multiple meanings all at once. Such mechanism predicts involvement of contextually inappropriate meanings and subsequently suppression of such meanings (Fodor 1983; Swinney 1979).

In the field of nonliteral language, Grice's (1975) *literal first model* – a lexically based model – assigns temporal priority to utterance literal interpretation, irrespective of contextual information. Adjusting this interpretation to contextual information is a later process, involving automatic suppression of inappropriate (literal) interpretations and consequential derivation of the appropriate (nonliteral) interpretations. Like *the direct access view*, then, *the literal first model* admits no inappropriate interpretations in the final output of the interpretation process.

According to another model, *the graded salience hypothesis* (see section 1.2 below), lexical access is exhaustive but ordered, making available salient – coded and prominent meanings – before less-salient (coded but less prominent) ones

---

2 One should note, however, that Campbell and Katz (2012) assume a weaker version of the model according to which "various linguistic and extralinguistic information provide probabilistic (but not necessary) support for or against a sarcastic interpretation" (p. 459).

(Giora 1997, Giora 2003; see also Duffy et al. 1988; Kawamoto 1993). And although some of the meanings and the interpretations constructed on the basis of these meanings must be contextually incompatible, their suppression is not automatic but sensitive to discourse goals.

1.2 Salience

What is meaning salience, then? According to *the graded salience hypothesis* (Giora 1997, Giora 1999, Giora 2003), meaning salience is a graded notion, ranging between salient, less-salient, and nonsalient outputs. Specifically,
–   a meaning is *salient* if it is coded in the mental lexicon and enjoys prominence due to cognitive factors (e.g., prototypicality) or experiential exposure (affected by e.g., familiarity, frequency, conventionality, etc.), regardless of degree of (non)literalness;
–   a meaning is *less-salient* if it is coded in the mental lexicon but is low on these factors (e.g., prototypicality, familiarity, frequency, etc.), regardless of degree of (non)literalness;
–   a meaning that is not coded is *nonsalient*; it is novel or derived, regardless of degree of (non)literalness. When it relates to a novel utterance, it results in either a salience-based or a nonsalient *interpretation*. (i) *Salience-based* interpretations are utterance interpretations <u>not</u> listed in the mental lexicon but constructed, based on the salient – coded and prominent – meanings of the utterance components, (Giora 1997, Giora 2003; Giora et al. 2007). (ii) *Nonsalient* interpretations are also utterance interpretations <u>not</u> listed in the mental lexicon. However, they are removed from the salient meanings of the utterance components (e.g., novel metaphors, novel ironies). *Nonsalience*, then, is also a matter of degree.

In terms of processing, salient and less-salient meanings get accessed automatically, regardless of contextual information or degree of (non)literalness. Still, access is sensitive to degree of salience – the more salient the meaning the faster it gets activated.

And since nonsalient meanings and interpretations are constructed rather than accessed, they are often slower to derive. Still, *salience-based* interpretations are activated unconditionally, whereas *nonsalient* interpretations rely more heavily on contextual information for their derivation. Given that salience-based interpretations are activated promptly, nonsalient interpretations might often involve incompatible salience-based interpretations initially. These can be literal, as the literal interpretation of some novel ironic utterances (*This knife is really*

sharp, said of a knife that wouldn't cut) or nonliteral, as the metaphorical interpretation of some novel ironic utterances (*This student is really sharp* said of a student who did extremely poorly on her entrance examination, see e.g., Colston and Gibbs 2002).

The graded salience hypothesis also allows contextual information to play a significant role in comprehension. It may affect processing initially via its predictive mechanism (Peleg et al. 2001, Peleg et al. 2004). It may also affect suppression of inappropriate interpretations once they are detrimental to the appropriate interpretation (Giora et al. 2007). However, even when strongly biasing, contextual information will not block salient meanings and hence salience-based interpretations when incompatible. According to *the graded salience hypothesis*, then, salient meanings and salience-based interpretations are privileged (Ariel 2002), enjoying priority even when incompatible.

Still, will these privileged meanings and interpretations be retained in memory even when incompatible? According to *the graded salience hypothesis*, incompatible meanings and interpretations may be retained if they are not detrimental to the interpretation process but instead play a role in supporting the intended meaning or interpretation. What role, then, could an incompatible meaning or interpretation play in constructing the contextually appropriate interpretation? A case in point is irony interpretation. According to *the indirect negation view*, the salience-based yet incompatible interpretation of an ironic utterance is instrumental in allowing the comprehender to weigh the gap between what is said and the reality that frustrates it (Giora 1995). According to the *relevance theoretic account*, assuming that irony projects a dissociative attitude from (the salience-based interpretation of) what is said while ridiculing it (Sperber and Wilson 1986), entails that this interpretation must be retained. According to *the tinge hypothesis*, the role of salience-based though incompatible interpretations is to mute the negative criticism of irony (Dews and Winner 1997, Dews and Winner 1999; Schwoebel et al. 2000). These are just examples indicating that salience-based interpretations do play a role in constructing the ironic interpretation despite their contextual inappropriateness. Their retention is therefore required, as predicted by *the retention/suppression hypothesis* (Giora 2003; Giora and Fein 1999a, Giora and Fein 1999b; Giora et al., 2007; Giora et al. 2014).

Indeed, the involvement of such "irrelevant" meanings and interpretations (e.g., the literal or metaphorical interpretation of novel, context-based sarcastic ironies) in the final output of utterance interpretation is quite prevalent (see e.g., Akimoto et al. 2012 with regard to intentional irony; Brisard et al. 2001; Burgers and Beukeboom in press; Colston and Gibbs 2002; Giora and Fein 1999b; Giora et al. 2007; Ivanko and Pexman 2003, [Experiment 3]; Pexman et al. 2000; Tartter et al. 2002).

1.3 Strong contextual support: The expectation hypothesis

Recall that according to context-based models, a strong context may activate contextually compatible meanings and interpretations immediately and exclusively even when nonsalient. What would count as a strong context, then? What are the factors that render contextual information strongly supportive of the appropriate interpretation so that it is facilitated immediately without having to go through the utterance inappropriate salience-based interpretation first? According to Gibbs (2002), such a context could be one that induces an expectation for the utterance in question. In the case of irony, a context that sets up an "ironic situation" through some contrast between what is expected by a protagonist and the reality that frustrates this expectation will induce an expectation for an ironic utterance on the part of the comprehender, which, in return, will facilitate the ironic interpretation (Gibbs 2002: 462).

Reading times studies, however, did not support this view. For instance, Giora et al. (2009) tested the "ironic situation" hypothesis by measuring reading times of identical statements (*Ski vacation is recommended for your health*) embedded in a context featuring an "ironic situation", in which the protagonist's expectation was frustrated, a "non-ironic situation", in which the protagonist's expectation was met, and a "no-expectation situation", which did not feature any expectation on the part of the protagonist (see examples a-c in Appendix). Results showed that target utterances took equally long to read following ironically biasing contexts, regardless of whether they made up an "ironic situation" or a "non-ironic situation"; they were, however, fastest to read when embedded in a "no-expectation context", which was biased toward the salience-based (here literal) interpretation (Giora et al. 2009 [Experiment 2]; see also Kreuz and Link, 2002, who show that expectation does not have an effect).

In fact, findings in Giora et al., (2009 [Experiment 1]) show that "an ironic situation" is not a strong enough context, as this kind of context did not favor an ironic utterance over a nonironic one. Rather, following both "ironic" as well as "non-ironic" situations, the salience-based statement was by far the more favored option (70%), while the ironic one fared poorly (30%). (For similar results, see also Ivanko and Pexman 2003).

So what kind of contextual information can induce an expectation for an ironic utterance? And will expecting it block salience-based interpretations? According to Pexman et al. (2000), speaker's occupation (e.g., one that is associated with nonliteralness, including irony) should facilitate irony interpretation, since such information should raise an expectation for a nonliteral (ironic) utterance on the part of the nonliteral (e.g., ironic) speaker. Results, however, show that

such contextual information affected longer rather than faster reading times of ironic statements compared to controls.

Giora et al. (2007), aimed to come up with contextual information shown to raise an expectation for an ironic utterance in order to find out whether such a strong context can indeed make a difference and facilitate ironic interpretations while blocking salience-based ones. Based on studies demonstrating that expectancy may be built up by the preceding stimulus sequences (Jentzsch and Sommer 2002; Kirby 1976; Laming 1968, Laming 1969; Soetens et al. 1985; see also Burgers et al. 2013), Giora et al., created an environment that would allow participants to anticipate an ironic utterance on the basis of prior occurrences of such stimuli, whether in specific contexts (Experiment 1) or via the manipulation of the experimental design (Experiments 3–4).

Giora et al.'s (2007) Experiment 1 showed that a prior occurrence of an ironic statement indeed induced an expectation for another such turn. Specifically, it showed that a context featuring an ironic speaker in context mid-position (examples d–e in Appendix; bold added) prompts participants to expect another such statement on the part of that speaker, which, in this case, was preferred over a salience-based, literal statement. Although both contexts were shown to raise an expectation for a sarcastic utterance, participants were faster to read the same target statement when biased toward the salience-based (here literal) interpretation than toward the context-based (here ironic) interpretation.

In Giora et al.'s (2007) Experiments 3–4, we further tested the resistance of salience-based interpretations to a strong context. Given that a prior occurrence of an ironic statement strengthens contextual expectation for another such turn (see Giora et al. 2007 [Experiment 1]), these experiments aimed to raise an expectation for an ironic utterance by manipulating multiple recurrences of ironic statements via the experimental design.

In these experiments participants were presented either texts that always ended in an ironic statement (+Expectation condition) or texts that were equally divided between ironic and nonironic endings (–Expectation condition). The measure was response times to probes related to either the appropriate non-salient ironic interpretation or to the inappropriate salience-based, nonironic interpretation. They were collected at two (750, 1000 ms) interstimulus intervals (ISIs). Four pretests controlled for (a) the similar salience status of the probes measured online outside a biasing context; (b) the probes' relatedness to the interpretation (rather than to the lexical meanings) of their respective targets – either the ironic or the nonironic; (c) the probes' reliance on the interpretation of the target sentence in its context rather than on the context alone; and (d) for the ironic bias of the ironically biasing contexts and the nonironic bias of the salience-based biasing context.

Findings showed no differences between these two conditions, whether at the short (750 ms) or long (1000 ms) delays. As predicted by *the graded salience hypothesis*, in both (+/−Expectation) conditions, salience-based probes were always faster to respond to compared to nonsalient ironically related ones, regardless of length of ISI.

Evidence so far, then, has failed to demonstrate that strong contextual information, inducing an expectation for an ironic utterance, can make a difference. Such contexts did not filter out incompatible salience-based interpretations, which surfaced initially; nor did they facilitate ironic interpretations immediately. Allowing comprehenders long enough processing time did not make a difference either (for a review, see Giora et al. 2007; for more recent corroborating results, see Filik et al. 2014; Filik and Moxey 2010).

However, given that our earlier studies have involved implicit contextual cues, in our studies here, we reinforce previous contextual support by adding contextual cues that are overt, explicitly alerting comprehenders as to an oncoming sarcastic utterance.

2 Reinforcing contextual support by explicit contextual cues

In his commentary on Giora et al. (2009), Katz (2009: 404) proposed that future research of irony should focus on reflecting the ecology of irony which "is rich in the number and strength of "hints" or "constraints" that invite irony". Indeed, in the two experiments reported here, we follow this suggestion, multiplying the number of contextual constraints supportive of an ironic interpretation, compared to those used in Giora et al. (2007).

In the two studies reported here, we used materials used in Giora et al. (2007). Here, however, items were presented in richer contexts, reinforced by multiplying the number of supportive contextual cues, which, in addition, were also explicit. Specifically, in Experiment 1, revised versions of dialogues used in Giora et al. (2007 [Experiment 1]) and some additional new dialogues were enriched by overt adverbial cues alerting comprehenders as to whether a **specific** utterance was intended ironically or nonironically; this was true of all the ironic and nonironic statements in dialogue mid and final position (except for 5 out of 22 salience-based biased items, in which the final target utterance was not marked by a cue). Prompting an expectation for an ironic interpretation of a specific utterance by cuing that very same utterance must be a strong if not a stronger "hint" than inducing an expectation for a nonspecific ironic utterance, as was the case in

Giora et al.'s (2007) manipulations. Indeed, as shown below (see Experiment 1, footnote 5), there was some increase in contextual strength.

Measures here were reading times of target utterances and lexical decisions to (ironically/salience-based) related, unrelated, and nonword probes at various ISIs (750, 1500, 2000 ms). Reading times were employed since they may indicate processing effort. However, since they may still be somewhat opaque with regard to when, why, and which interpretation affects this difficulty, lexical decisions were employed too. And various ISIs were used in order to find out whether additional processing time might be informative as to whether suppression of salience-based interpretations would take place. As in Giora et al. (2007), we aimed to show that, regardless of strength of contextual support to the contrary, salience-based interpretations of ironic utterances would not be preempted, nor would they be suppressed[3], as predicted by the *graded salience-hypothesis* (Giora 1997, Giora 2003).

In Experiment 2 here, we aimed at replicating previous results (Giora et al. 2007 [Experiments 3–4]) by strengthening the context via adding an explicit cue. In Giora et al.'s (2007 [Experiments 3–4]), the fact that we studied irony interpretation might have been implicitly conveyed to participants (in the +Expectation condition). Here, in this same condition, we made it explicit, by informing participants that the experiment tested sarcastic irony interpretation. Additionally, we employed here a larger range of processing times. Shorter (750 and 1000 ms ISIs) and longer intervals (1500, 2000, 2500, and 3000 ms ISIs) would allow us an insight into both initial activation processes and late suppression/retention processes. In sum, the contexts and experimental design used in Experiments 1–2 proliferate ironic uses and ironic cues within (Experiment 1) and between (Experiment 2) items, at various processing stages. Predictions tested here remain the same, as before. Weighing salience-based effects against contextual strength effects under stricter conditions should allow us to test the predictions of the expectation-based accounts vis à vis those of the graded salience hypothesis.

3 Experiment 1

To contrast predictions of *the expectation hypothesis* (Gibbs 2002) with those of *the graded salience hypothesis* (Giora et al. 2007) under stricter conditions, we have revised dialogues used in Giora et al. (2007) and also increased the number of items. In addition, here, both types of context were revised so as to be similarly

---

3 At best, such interpretations can decay over time.

strongly constraining in terms of the type of utterance they could give rise to. As in Giora et al. (2007), the ironically biasing contexts used here include an ironic utterance in dialogue mid-position in addition to an ironic target in dialogue final position. However, to further strengthen the expectation for an ironic utterance, these 2 ironic utterances are now preceded by an ironic cue (such as "derisively"), which discloses the speaker's ironic intent. Along the same lines, the salience-based biasing contexts now feature only salience-based utterances. In addition, 2 utterances, including the target utterance, are now preceded by a cue (such as "sadly"), which is revealing of the speaker's nonironic intent.

On the basis of an established similar anticipatory strength of both contexts (inviting either a sarcastic or a nonsarcastic utterance), *the expectation hypothesis* would predict equal processing times of identical targets (embedded in sarcastically and salience-based biasing contexts) and equally faster response times to related (sarcastic/salience-based) probe-words in their respective contexts compared to unrelated ones. In contrast, according to *the graded salience hypothesis*, salience-based interpretations should be facilitated faster than nonsalient (here ironic) ones, regardless of strength of contextual bias.

3.1 Methodology

Design. A $3 \times 2 \times 3$ factorial design was used, with ISI (750 ms, 1500 ms and 2000 ms) as between-participants factor, and context type (ironically-biasing vs. salience-based-biasing) and probe type (ironically-related, salience-based-related, and unrelated) as within-participants factors.

Participants. Participants in Experiment 1 were 72 students (55 women and 17 men), between the ages of 19 and 33. All participants were students of Yezreel Valley College who volunteered to take part in the experiments. Twenty-four participants were assigned to each of the 3 ISI conditions.

Materials. Materials included 22 pairs of Hebrew dialogues, 10 utterances long, which took place between friends (1–2 below). Target utterances for each pair were identical and appeared in texts' final position. For each pair, one dialogue context biased the target utterance toward its salience-based (often literal) interpretation, while its counterpart biased it toward the nonsalient, ironic interpretation. For the ironically biasing dialogues, the 6th utterance was always a sarcastic irony (bold added, for convenience) and so was the final utterance – the target sentence – (bold added, for convenience), spoken by the same ironic speaker (2 below). The dialogues biasing toward the salience-based interpretation were very similar to the ironically biasing dialogues both in structure, length, and content,

except for the salience-based bias of the 6th and final target utterances and the absence of an ironic speaker (1 below). Materials further included 20 equally long filler dialogues ending in either a literal or a metaphorical utterance. Following each of the items, 44 yes/no comprehension questions were displayed. In addition, for all target utterances, two types of probes were prepared, either (salience-based/ironically) related or unrelated to their interpretation; nonwords were presented at offset of filler texts only:

(1) *Dani (rubbing his stomach): Do you have anything to eat around here?*
    Iris: Want me to make you a sandwich?
    Dani: I'd like a proper meal, I'm starving.
    Iris: Haven't you eaten anything today?
    Dani: Not really. A few snacks.
    Iris (worrying): **You really should be more careful about what you eat**.
    Dani: Don't worry. Today was not a typical day. Usually I am very strict with myself.
    Iris: I'm happy to hear that.
    Dani: I go to the gym three times a week, and eat only low fat foods.
    Iris (impressed): **I see you've developed some great habits**.

(2) *Dani (rubbing his stomach): Do you have anything to eat around here?*
    Iris: Want me to make you a sandwich?
    Dani: I'd like a proper meal, I'm starving.
    Iris: Haven't you eaten anything today?
    Dani: I've had a couple of chocolate bars and two donuts.
    Iris (winking): **I see you're on a strict diet**.
    Dani: Since I quit smoking I'm gobbling sweets all the time.
    Iris: I didn't even know you used to smoke.
    Dani: Well, I started smoking so I could quit sniffing glue.
    Iris (mocking): **I see you've developed some great habits**.
    Probes: Salience-based – healthy; Ironic – harmful; Unrelated – fragile.

Three pretests controlled for the ironic bias of the ironically biasing dialogues, which induced a significantly stronger expectation for an ironic utterance compared to the nonironic dialogues (pretest 1); the similar salience status of the 3 types of probe-words, which were measured online in terms of response times, following neutral contexts (pretest 2); and the equivalent relatedness of the related probes to the interpretation of their relevant target utterances in their respective contexts and the unrelatedness of the unrelated probes (pretest 3).

Pretest 1: Controlling for the ironic bias of the ironically biasing dialogues.
To ascertain that the ironically biasing dialogues indeed induce an expectation for an ironic utterance in final position compared to salience-based biased dialogues, we presented two groups of 15 participants – in all, 30 volunteers, first and second year students of Linguistics at Tel Aviv University, with two types of booklets, each containing the 22 experimental items (half biased toward the ironic interpretation and half – toward the salience-based interpretation), presented without the target sentence but including the biasing cues. In addition, 20 filler dialogues were also included. Each dialogue was followed by two endings presented in a pseudo-random order: the ironically biasing contexts were followed by the ironic target (*I see you've developed some great habits*) and a corresponding salience-based ending (*With such habits you might damage your health*). Similarly, the salience-based biasing contexts were followed by the salience-based target (*I see you've developed some great habits*) and a corresponding ironic ending (*With such habits you might damage your health*). Participants were asked which of the two alternatives they would expect to end the dialogue. Findings showed that, following dialogues featuring an ironic speaker as well as ironic cues, participants selected an ironic ending in 13.5 (*SD* = 1.5) out of the 15 cases. However, following dialogues not featuring an ironic speaker or ironic cues, but featuring, instead, nonironic speakers and cues (or no cues), participants selected an ironic ending only in 1.5 (*SD* = 1.8) out of the 15 cases, $t_1(29) = 19.97$, $p < 0.001$, d = 3.65; $t_2(21) = 32.12$, $p < .001$, d = 6.85.[4] These findings demonstrate that the dialogues featuring an ironic speaker and ironic cues induced a much stronger expectation for an oncoming ironic utterance compared to when no such cues or speakers were provided (and also compared to those in Giora et al. 2007).[5] These findings further establish that contextual bias toward both, the ironic and the salience-based interpretations, was identical in strength.

Pretest 2: Controlling for the similar salience status of the 3 types of probe-words.
To ascertain that the related and unrelated probes are of similar salience status, 33 students of Yezreel Valley College (27 women and 6 men), between the ages of 22 and 29, participated in an online pretest for a course credit. They were seated

---

4 Here "cases" refers to the number of participants that selected the ironic ending in each item, averaged across items in each type of dialogue (i.e., featuring and not featuring an ironic speaker).
5 There is also a slight increase in selecting ironic endings here compared to Giora et al. (2007) Experiment 1. Whereas in Giora et al. (2007) ironic endings were selected in 12.2 out of 16 (76%) cases, here they were selected in 13.5 out of 15 (90%) of the cases.

in front of a computer monitor and were asked to make lexical decisions as to whether a letter string, presented at a short ISI (250 ms), following offset of a neutral sentence-context, made up a Hebrew word or not. Materials were displayed centrally and included 46 neutral sentence-contexts and 4 probes: salience-based related, ironically related, unrelated, and nonword probes. Of the sentence-contexts, 22 were followed by the experimental word probes and 24 were followed by the nonword probes.

Participants self-paced their reading of the neutral sentence-contexts and then made a lexical decision by pressing a "yes" or a "no" key. They were given 16 practice trials. Results exhibited no difference between the salience-based related (803 ms, $SD = 287$), ironically related (812 ms, $SD = 219$), and unrelated (810 ms, $SD = 277$) probes. This was confirmed by both participant and item ANOVAs, $F_1(2,64) < 1$, n.s., $F_2(2,42) < 1$, n.s. We thus guaranteed that the experimental probes were of comparable salience (i.e., accessibility).

Pretest 3: *Controlling for the equivalent relatedness of the related probes.*
To make sure that the salience-based and ironically related probes were similarly related to the interpretation of their relevant target utterances in their respective contexts, and that the unrelated probes were indeed unrelated, we ran another pretest. In this pretest, 20 volunteers, first and second year students of Linguistics at Tel Aviv University, were presented the experimental texts, each followed by three probes words – the related (whether salience-based or ironic), the unrelated, and an additional semi-related word, which was used only to allow a choice of mid-position on the scale. Participants were asked to rate each word on a 7-point relatedness scale. Specifically, they were asked to rate the extent to which the probe word was related to the interpretation of the last sentence in its given context.

Results showed that, in the ironically biasing contexts, the ironically related word was highly related (5.12, $SD = 0.97$), while the unrelated word was indeed unrelated (2.20, $SD = 1.09$). Similarly, in the salience-based biasing contexts, the salience-based related word was highly related (5.30, $SD = 1.05$), while the unrelated word was unrelated (2.27, $SD = 0.97$). Two 2×2 ANOVAs with context type (ironic, salience-based) and relatedness (related, unrelated) as factors were run, one for participants and one for items. They revealed no effect of context type and no interaction, only a relatedness effect, $F_1(1,19) = 81.6$, $p < .001$, partial $\eta^2 = .81$; $F_2(1,21) = 306.1$, $p < .001$, partial $\eta^2 = .93$, as expected. These results thus guaranteed that the salience-based and ironically related probes were similarly related to the interpretation of the experimental items and targets they were designed to test. They further guaranteed that the unrelated probes were indeed unrelated.

Procedure. Participants were told the investigators were interested in comprehension and were instructed to read the texts, make lexical decisions, and respond to a comprehension question. They were seated in front of a computer monitor and were given 16 trial sessions involving initially only a lexical decision task followed by 3 full trial sessions (including dialogues, lexical decisions, and comprehension questions). Then they self-paced their reading of the experimental texts, advancing them sentence by sentence. Specifically, the first key press displayed the speaker's name and the ironic cue when present, and the next key press displayed the utterance. Following target utterances, participants made a lexical decision to a letter string displayed centrally at either 750, 1500, or 2000 ms ISI. The pressing of the key further displayed a yes/no comprehension question. Reading times of target sentences and response times to probes were measured by the computer.

3.2 Results

Reading times. Reading times larger than 2 SDs above the mean of each participant were considered outliers and were excluded from the analyses. Overall, 92 reading times, out of 1584, were excluded (5.8%). Mean reading times of the target sentences were used as the basic data for the analyses. Results are presented in Table 1.[6]

A two-way ANOVA was performed for both participant (F_1) and item (F_2) analyses. Those 3×2 ANOVAs (with ISI and context type as factors) showed a significant ISI effect, $F_1(2,69) = 6.56$, $p < .005$, partial $\eta^2 = .16$; $F_2(2,42) = 87.67$, $p < .001$, partial $\eta^2 = .81$, and, more importantly, a significant context type effect, $F_1(1,69) = 14.74$, $p < .001$, partial $\eta^2 = .18$; $F_2(1,22) = 9.39$, $p < .005$, partial $\eta^2 = .31$,

Table 1: Mean Reading Times: Experiment 1 (SD in Parentheses).

ISI	Context Type	
	Ironic Context	Literal Context
750 ms	1716 (620)	1520 (454)
1500 ms	1495 (421)	1352 (374)
2000 ms	1217 (471)	1124 (358)
All ISIs	1476 (544)	1332 (425)

---

6 In Tables and Figures we use "Literal" for "Salience-based".

Table 2: Mean Response Times to Probes: Experiment 1 (SD in Parentheses).

ISI	Context Type	Probe Type			
		Ironically Related	Literally Related	Unrelated	Overall Mean
750 ms	Ironic context	873 (231)	840 (280)	860 (222)	858
	Literal context	873 (220)	793 (213)	814 (187)	827
1500 ms	Ironic context	744 (162)	732 (170)	755 (161)	744
	Literal context	751 (177)	771 (196)	771 (199)	764
2000 ms	Ironic context	809 (255)	773 (191)	809 (276)	797
	Literal context	810 (269)	775 (198)	830 (229)	805
All ISIs	Ironic context	809 (223)	782 (220)	808 (226)	799
	Literal context	811 (228)	780 (200)	805 (204)	799

but no interaction (F < 1). As predicted by the *graded salience hypothesis*, salience-based targets were faster to read (1332 ms; $SD = 425$ ms) than equally highly expected nonsalient ironic targets (1476 ms; $SD = 544$ ms).

Response times. Response times larger than 2 SDs above the mean of each partic-ipant were considered outliers and were excluded from the analyses. Overall, 67 response times, out of 1584, were excluded (4.2%). Mean response times to ironically-related, salience-based-related, and unrelated probes were used as the basic data for the analyses. Results are presented in Table 2.

A three-way ANOVA was performed for both participant (F_1) and item (F_2) analyses, with ISI as a between-participant factor and context type (ironically-biasing vs. salience-based-biasing) and probe type (ironically-related, salience-based-related, and unrelated) as within-participants factors. Those $3 \times 2 \times 3$ ANOVAs showed no context type effect or interactions with context type (Fs < 1). However, there was a significant probe type effect in the subject analysis, $F_1(2,138) = 3.42$, $p < .05$, partial $\eta^2 = .05$; $F_2(2,42) = 2.06$, $p = .14$, partial $\eta^2 = .09$. When two repeated contrasts between probe-types were conducted, the Salience-based – Ironic comparison was significant, $F_1(1,69) = 4.88$, $p < .05$, partial $\eta^2 = .07$; $F_2(1,21) = 4.32$, $p < .05$, partial $\eta^2 = .17$, while the Salience-based-Unrelated com-parison was marginally significant in the subject analysis, $F_1(1,69) = 3.88$, $p = .053$, partial $\eta^2 = .05$; $F_2(1,21) = 2.51$, $p = .128$ partial $\eta^2 = .11$. The lack of context type × probe type interaction suggests that response facilitation found for salience-based-related probes was equally strong following ironically and salience-based biasing contexts, as predicted by the *graded salience hypothesis* (see Figure 1).
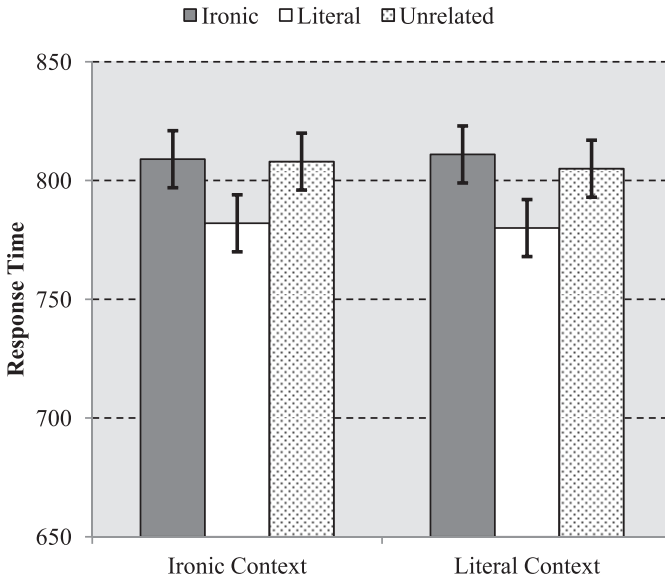
Fig. 1: Mean response times to probes in all ISIs. Error bars represent standard errors. [Standard errors in all figures were calculated according to Loftus and Mason's (1994) recommendations for within-subject designs.].

3.3 Discussion

In Experiment 1 we weighed *the graded salience hypothesis* (Giora 2003; Giora et al. 2007) against *the expectation hypothesis* (Gibbs 2002). According to the latter, a strong context inducing an expectation for an ironic utterance will facilitate that interpretation immediately and directly; according to the former, a strong context cannot block salient meanings and hence salience-based interpretations which will be activated quite early on in the interpretation process. Testing the predictions of these hypotheses was the aim of this experiment, which was geared toward replicating the results of Giora et al.'s (2007) Experiment 1, only under stricter conditions. In Giora et al.'s (2007) Experiment 1, reading times of targets, embedded in salience-based biasing contexts, were faster than those of their counterparts, embedded in ironically biasing contexts.

In Experiment 1 here, we used similar dialogues but strengthened contextual information even further (as would be required by *constraint satisfaction models*, see Katz 2009 and Campbell and Katz 2012). The strength of the ironically biasing dialogues, which featured an ironic speaker in mid-position, was amplified by adding explicit information about the intent of that speaker who also uttered the

target sentence; similarly, the strength of the salience-based biasing dialogues, which involved only literal speakers, was amplified by adding explicit information about the salience-based intent of the speaker who also uttered the salience-based target sentence. To test the processing routes of these biased targets, we employed two measures: reading times of targets (as in Giora et al., 2007) and lexical decisions to related and unrelated probes at 3 different ISIs (750, 1500, 2000 ms).

Despite strengthening expectancy for an ironic utterance, results replicated previous findings.[7] They attest to the temporal priority of (equally strongly biased) salience-based interpretations: salience-based (nonironic) utterances were always faster to read compared their nonsalient ironic counterparts.

Such results argue against *the expectation hypothesis*, which predicts equivalent reading times for nonsalient sarcastic ironies in contexts inducing an expectation for an ironic utterance compared to salience-based interpretations in contexts inducing an expectation for a salience-based interpretation. Instead, while replicating previous results (e.g., Colston and Gibbs 2002; Filik et al. 2013; Filik and Moxey 2010; Giora and Fein 1999a; Giora et al. 1998; Giora et al. 2007; Pexman et al. 2000; Schwoebel et al. 2000), these findings argue in favor of the view that context cannot bypass salient although inappropriate meanings and hence salience-based yet inappropriate interpretations. The latter, therefore, emerged earlier than nonsalient yet context-based interpretations, on account of their reliance on salient meanings.

Results from lexical decisions corroborated the reading times results. They show that response times to salience-based related probes were facilitated earlier than ironically related probes, which were not facilitated. Such results support *the graded salience hypothesis* which predicts the priority of salience-based interpretations which will not be blocked by a strong context. These results further argue against the suppression of salience-based "irrelevant" interpretations (see Grice 1975). Instead, they attest to the retention of such interpretations which are instrumental in the construction of the appropriate interpretations (see Giora 2003).

In the following experiment, we further weigh *the graded salience hypothesis* (Giora et al. 2007) against *the expectation hypothesis* (Gibbs 2002) and *the constraint-based model* (Campbell and Katz 2012; Katz 2009; Pexman et al. 2000)

---

7 Note that in Giora et al. (2007), the salience-based literally biasing contexts were weaker than those here in that they included an ironic speaker in dialogue mid-position (which prompted a preference for an ironic interpretation of the last target utterance). Still, salience-based interpretations were not slowed down. Rather, they were faster than the ironic ones, derived in the ironically biasing contexts.

by further increasing the number of constraints used in Giora et al. (2007 [Experiments 3–4]) and by making them more explicit. In addition, longer processing times were allowed by increasing the length of ISIs to see whether additional processing time might afford suppression of salience-based interpretations (as predicted by Grice 1975). As before, we anticipate replication of these previous results which showed both facilitation and retention of salience-based (often literally) related probes.

4 Experiment 2

Recall that in Giora at al.'s (2007) Experiment 3–4, we manipulated an expectation for an ironic utterance via the design of the experiment. Thus, in one condition – the +Expectation condition – participants were presented items all of which ended in an ironic utterance; in the other – the –Expectation condition – participants were presented items half of which ended in a salience-based utterance and half in an ironic utterance. Results showed no expectancy effects on sarcastic irony interpretation. Specifically, no differences were found between the +/–Expectation conditions. Instead, in both conditions, and regardless of length of ISI (750, 1000 ms), the pattern of results was the same: only probes related to the salience-based interpretation were facilitated.

In Experiment 2 here we aim to replicate these results under stricter expectancy conditions. While the design of the experiment is as in Giora et al.'s (2007) Experiment 3, the +Expectation condition here is strengthened by an explicit cue: participants were told that the experiment tested sarcastic irony interpretation. In addition, we used 6 different ISIs here – 750, 1000, 1500, 2000, 2500 and 3000 ms – which should allow testing suppression of inappropriate interpretations, when and if it takes effect.

4.1 Methodology

Design. A 6 × 2 × 3 factorial design was used with ISI (750, 1000, 1500, 2000, 2500 and 3000 ms), and expectancy (+/–Expectation) as a between-participants factor and word type (ironically/salience-based related and unrelated) as a within-participants factor.

Participants. Participants were 432 students of Tel-Aviv University and their friends, between the ages of 16 and 42 (72 participants for each ISI condition). They were paid 30 shekels (approx. $7) for their participation.

Materials. Materials are as in Experiment 3 in Giora et al. (2007) (see 3–4 below, bold added), which were controlled for similar contextual bias, similar relatedness of probes to the targets in their respective contexts compared to the unrelatedness of the unrelated probes, the probes' salience status, and the similar (ironic/salience-based) bias of the respective contexts. However, since attempting to control for the similar salience status of the probe-words revealed salience differences between the salience-based related, the ironically related, and the unrelated probes, these results served as baseline means:

(3) *John was a basketball coach. For the past week he was feeling restless, worrying about the upcoming game. It was yet unclear how the two teams matched up, and he was anxious even on the day of the game. When he got a call telling him that the three lead players on the opposing team will not be able to play that night, John wiped the sweat off of his forehead and said to his friend: "**This is terrific news!**"*

(4) *John was a basketball coach. For the past week he was feeling restless, worrying about the upcoming game. It was yet unclear how the two teams matched up, and he was anxious even on the day of the game. When he got a call telling him that the three lead players on his team will not be able to play that night, John wiped the sweat off of his forehead and said to his friend: "**This is terrific news!**"*
Probes: Salience-based related – winning; ironically related – losses; unrelated – friends.

As in Giora et al. (2007), materials here were divided into two subsets, each comprising two blocks of items – one inducing a specific expectation for an ironic target and treated as "+Expectation" condition and one not inducing a specific expectation and treated as "–Expectation" condition. One block – the set inducing an expectation for a sarcastic irony – comprised 36 contexts, all biased toward their ironic interpretation. This block included 3 ironic buffers, 15 experimental ironic contexts, and another 18 ironic filler contexts. The other block – the set not inducing a specific expectation – also comprised 36 contexts. They included the same 18 contexts appearing in the other block – 15 experimental ironic contexts and 3 ironic buffers – and another 18 filler contexts, biasing their final statement toward their salience-based (here literal) interpretation. The items were arranged in four versions, and, with the exception of the buffers, were presented pseudo-randomly. In this way, half of the participants received half of the ironies as experimental contexts and the rest of the participants received the other half of the ironies as experimental contexts. Each time, the ironies that were not

used as experimental items were used as expectancy generators to form the "+Expectation" condition, whereas the salience-based contexts were used to create the "−Expectation" condition.

Procedure. Participants were presented experimental texts which were displayed centrally. Only participants in the "+Expectation" condition were told in advance that sarcastic irony interpretation was being examined. They were asked to read the items and advance the contexts sentence by sentence at their own pace. The last target sentence was advanced word by word. Following the reading phase, participants were asked to make a lexical decision to a letter string displayed centrally after an ISI of either 750, 1000, 1500, 2000, 2500 or 3000 ms. The pressing of the key further displayed a yes/no comprehension question, 12 in all, which appeared in 33% of the cases.

4.2 Results

Response times larger than 2 SDs above the mean of each participant were considered outliers and were excluded from the analyses. Overall, 330 response times out of 6480, were excluded (5.1%). Mean response times were used as the basic data for the analyses. Results are presented in Table 3 and Figure 2.

Since the items we used exhibited salience differences in a salience pretest we had run earlier (see Giora et al., 2007, Experiment 3), we decided to take into account the baseline for each item found in our pretest. Therefore, we conducted only item analysis, where subtraction of these baselines was allowed. The findings presented here are the results after this subtraction.

A $6 \times 2 \times 3$ ANOVA (with ISI, expectancy, and probe type as factors) showed a significant probe-type effect, $F(2,58) = 6.05$, $p < .005$, partial $\eta^2 = .17$. To determine where the differences in the probe-type factor originated, two repeated contrasts were conducted. Results showed that both the Salience-based – Ironic comparison and the Salience-based-Unrelated comparison were significant, $F(1,29) = 14.45$, $p < .005$, partial $\eta^2 = .33$, $F(1,29) = 5.36$, $p < .05$, partial $\eta^2 = .16$. As can be seen in Table 3, salience-based related probes were faster to respond to than ironically related probes. Note further that there was no ISI × probe type interaction ($F(10,290) = 1.24$, p = .27), suggesting that this trend holds for all ISIs.

Most importantly, although there were also significant main effects of ISI and expectancy ($F(5,145) = 52.07$, p < .001, partial $\eta^2 = .64$, and $F(1,29) = 31.00$, p < .001, partial $\eta^2 = .52$, respectively), there was no expectancy × probe type interaction, $F(2,58) = 1.14$, p = .33. Namely, expectation did not have a differential effect on the three types of probes (see Figure 2). The only other significant effect was the ISI × expectancy interaction, $F(5,145) = 4.62$, p < .005, partial $\eta^2 = .14$.

Table 3: Mean Response Times to Probes (after subtraction of baseline means): Experiment 2 (SD in Parentheses).

ISI	Expectancy	Probe Type			
		Ironically Related	Literally Related	Unrelated	Overall Mean
750 ms	−Expectation	295 (260)	240 (224)	275 (221)	270
	+Expectation	337 (189)	182 (149)	233 (160)	251
1000 ms	−Expectation	378 (213)	246 (190)	256 (213)	293
	+Expectation	341 (256)	183 (194)	296 (237)	273
1500 ms	−Expectation	185 (104)	163 (417)	189 (291)	179
	+Expectation	276 (249)	241 (316)	249 (248)	255
2000 ms	−Expectation	231 (225)	148 (158)	234 (188)	204
	+Expectation	152 (184)	60 (176)	170 (150)	128
2500 ms	−Expectation	170 (210)	110 (173)	147 (220)	142
	+Expectation	54 (146)	−42 (122)	46 (143)	19
3000 ms	−Expectation	143 (149)	72 (145)	126 (157)	113
	+Expectation	97 (148)	4 (144)	81 (186)	61
All ISIs	−Expectation	234 (112)	163 (122)	204 (124)	200
	+Expectation	210 (129)	105 (104)	179 (138)	165

4.3 Discussion

In this experiment, we test *the expectation hypothesis* (Gibbs 2002) and the *suppression hypothesis* (Grice 1975). We weigh them against the *graded salience hypothesis* (Giora 2003; Giora et al. 2007). To test *the expectation hypothesis,* an expectation for an ironic utterance was induced in two ways. First, as in Giora et al. (2007), it was implicitly prompted via the design of the experiment. Thus, in one condition – the +Expectation condition – participants were exposed to items all of which ended in an ironic utterance; in the other – the −Expectation condition – participants were exposed to items half of which ended in an ironic utterance and half – in a salience-based biased utterance. Second, to further strengthen expectancy, an explicit cue was added to the +Expectation condition only: participants in this condition were told that sarcastic irony interpretation was tested. To test *the suppression hypothesis* we allowed a wide range of processing times so that conditions varied between 750, 1000, 1500, 2000, 2500 and 3000 ms ISIs. Such a wide range should allow us to observe if and when suppression processes might kick in.
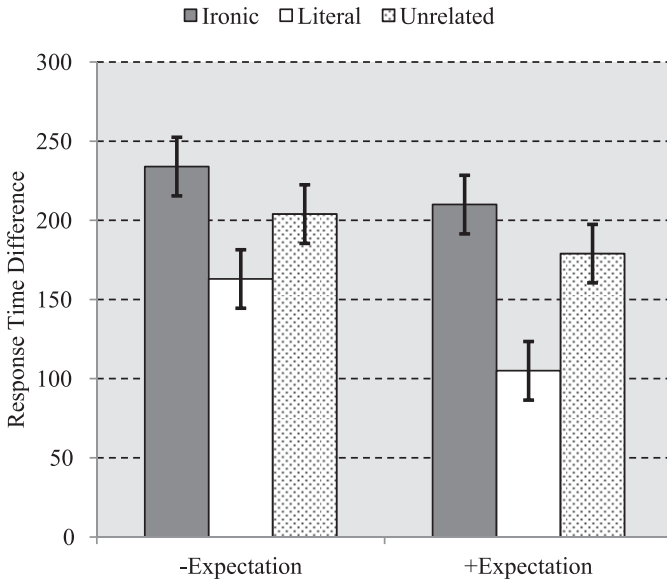
Fig. 2: Mean response times in all ISIs (after subtraction of baseline means). Error bars represent standard errors.

Although contextual expectation for a nonsalient ironic interpretation was evidently strong, it did not affect the patterns of results. In all 6 conditions, salience-based related probes were always faster to respond to than ironically related probes, as is evident by the significant Salience-based – Ironic comparison. The lack of ISI × probe type interaction suggests that this trend holds for all ISIs.[8]

As can be seen in Table 3, overall, it took less time to respond to probes in the +Expectation condition.[9] This cannot be accounted for by any of the theories tested here, and is irrelevant to them. The important point, however, is that there was no expectancy × probe type interaction. Recall that *the expectation hypothesis* (Gibbs 2002) and *the constraint-based model* of irony interpretation (Katz 2009; Campbell and Katz 2012; Pexman et al. 2000) predict initial facilitation of targets prompted by multiple cues inducing an expectation for an ironic

8 Indeed, the Salience-based – Ironic comparison was significant in each ISI condition, when analyzed separately.
9 Note, however, that it was not the case in the 1500 ms ISI condition, which probably caused the expectancy × ISI interaction.

utterance. Therefore, an expectancy × probe type interaction (expectancy having an effect on the ironically-related probe) is crucial for their models.

As for the suppression hypothesis (Grice 1975), there is no evidence suggesting suppression of salience-based albeit incompatible interpretations, given the significant Salience-based-Unrelated comparison.

In all, our results argue in favor of *the graded salience hypothesis* which predicts that salient meanings and salience-based interpretations will not be preempted by contextual information. Recall, however, that *the graded salience hypothesis* has no specific predictions as to when nonsalient (ironic) interpretations might be activated. It only maintains that salience-based interpretations will not be blocked, and, in the case of irony, they will not be suppressed either.

Still, one might wonder why our findings do not attest to irony facilitation even when long (3000 ms) processing time is allowed. Could it be the case that the ironic cues we used were not effective enough? The results of the various pretests answer this question in the negative. They ascertained that the contexts used were strongly biased towards the ironic interpretation. However, it is quite possible that they only facilitated irony detection.[10] Moving on to the next stage, where retrieving the ironic interpretation (and its implications) takes place, proves to be a more complex process. Indeed, in this study we examine **utterance interpretations** and their implications/implicatures in a given context (rather than lexical **meanings** of utterance key words). Probe words are therefore not directly related, not even in a contrastive manner, to the coded meaning of target's key words, especially when it comes to irony interpretation.

For instance, in the salience-based biased context (3), *terrific* is related to the implicated "winning" rather than only to its nonliteral encoded meaning – "great"; the sarcastically biased probe (in 4), however, should involve a longer chain of inferences. It should be interpreted as related to the implicated "losses" rather than only to "great" and then to "bad" (the inferred opposite of the salient "great" meaning of *terrific*, which should also be activated). The ironic interpretation then involves more processing steps than the salience-based interpretation, which are, in addition, more complex, and might therefore require much longer

---

10 One could interpret Regel, Coulson & Gunter's (2010) ERP evidence, showing early (and even late) brain sensitivity to speaker's style, along these lines. Thus, it is quite possible that sufficient exposure to pragmatic knowledge about speakers allows comprehenders to anticipate and detect ironic and literal statements, but not necessarily to construct or infer the ironic (or even the literal) interpretation and the implicatures involved. In addition, see Luka and van Petten (2014) on different contextual effects induced when different measures (such as ERPs and lexical decisions) are used.

interpretation time. This relation of the probes to utterance interpretations and implications in their respective contexts is what distinguishes them from the probes in our previous studies (Giora and Fein 1999a; Giora et al. 1998) where facilitation for irony was visible earlier than here. (On nonsalient ironic interpretations understood directly, faster than alternative, salience-based interpretations, see Giora et al. 2015; Giora et al. 2013).

5 General discussion

The focus of this study is on **constructed** interpretations of utterances (rather than on lexicalized meanings of utterances' components, retrieved directly from the mental lexicon). Three theories have been put to the test here, which differ in their predictions with regard to early and late interpretation processes. According to the *constraint-based* and *direct access* views, strong contextual information, rich in supportive constraints, will govern appropriate interpretation immediately and directly. Under such conditions, then, activating the appropriate interpretation will involve no inappropriate interpretations initially and therefore no suppressive processes later on (Gibbs 1986, Gibbs 1994, Gibbs 2002; Pexman et al. 2000).

According to the *literal first model* (Grice 1975), an utterance literal interpretation will be made available first, regardless of contextual information to the contrary. When a nonliteral interpretation is invited, adjusting the literal interpretation to contextual information is a late and effortful process, involving suppression of inappropriate interpretations.

According to *the graded salience hypothesis*, salience-based interpretations – interpretations constructed on the basis of the salient meanings of the utterance components – will not be blocked by a context strongly biased in favor of a nonsalient alternative. In addition, such inappropriate interpretations should not be discarded from the mental representation unless they are detrimental to the construction of the appropriate interpretation. In irony interpretation, such "irrelevant" interpretations are conducive to the utterance interpretation and will therefore be retained (Giora 2003; Giora et al. 2007; Giora et al. 2014).

Given that nonsalient irony involves irrelevant interpretations (e.g., its literal interpretation), examining its derivation could allow us to tease apart these three conflicting models. In two experiments we tested these views. Measures were reading times and response times. While reading times of an utterance can disclose interpretation difficulties in one context compared to another, response times to probes can be further revealing as to which interpretation was activated initially. Note that probes in this lexical decision task were related to the

interpretation ("winning"/"losses") of the target utterance ("This is *terrific* news!") in its specific context (game) rather than to the coded **meaning** ("great") of critical words (*terrific*) which must have been adjusted to contextual information. We aimed to find out whether a strong context might have an immediate effect, availing appropriate nonsalient interpretations directly and exclusively as proposed by the *direct access view* (Gibbs 2002).

Alternatively, would salience-based interpretations spring to mind initially, irrespective of contextual information to the contrary, as proposed by *the graded salience hypothesis* (Giora 1997, Giora 2003; Giora et al. 2007)? And if salience-based interpretations are made available initially, will they be suppressed later on (as argued by Grice 1975), or will they be retained since they are conducive (or at least non-intrusive) to the interpretation process (as proposed by Giora 2003; Giora et al. 2007).

Findings from the two experiments reported here tip the balance in favor of the temporal priority of salience-based incompatible interpretations over nonsalient compatible ones, while further arguing against their dismissal as irrelevant. In experiment 1, contexts made manifest an expectation for an ironic utterance by introducing an ironic speaker in dialogue midposition who also uttered the ironic target in dialogue-final position. The ironic intent of these utterances was further reinforced by means of an explicit ironic cue preceding this speaker's utterances (e.g., *mocking*), alerting comprehenders as to the ironic interpretation of a specific oncoming utterance. In the alternative nonironic versions, dialogues included no ironic speakers, and utterances were prefaced by salience-based biasing cues (e.g., *appreciating*). Measures were reading times and response times to related and unrelated (and nonword) probes (varying in terms of the processing times allowed, ranging between 750, 1500, and 2000 ms ISIs).

In Experiment 2, raising an expectation for an ironic utterance was manipulated in two different ways. First, expectancy was implicitly manipulated by exclusively exposing participants to items all of which ended in an ironic utterance (the +Expectation condition). Second, in this condition, participants were explicitly informed in advance that the experiment was testing sarcastic irony interpretation. In contrast, in the condition in which an expectation was not induced (the –Expectation condition), participants were exposed to items ending either in an ironic or in a nonironic utterance. In this condition, participants were not informed that irony interpretation was being tested.

Measures in this experiment were response times to related and unrelated (and nonword) probes, varying in terms of the processing times allowed, ranging between 750, 1000, 1500, 2000, 2500 and 3000 ms ISIs). Such a range should allow an insight into both initial activation and later suppression processes.

Results of these experiments show that, as predicted by *the graded salience hypothesis* (Giora 1997, Giora 1999, Giora 2003; Giora et al. 2007), salience-based interpretations were not inhibited. Instead, they were always activated initially, regardless of strength of contextual expectation for an ironic utterance. In fact, expectation for an ironic utterance did not affect the patterns of results which did not distinguish one condition (+Expectation) from another (−Expectation). Rather, both conditions attested to the temporal priority of salience-based interpretations over nonsalient (ironic) ones. They further suggested retention rather than suppression of salience-based although incompatible interpretations, as predicted by *the graded salience hypothesis*.

In addition to replicating Giora et al.'s (2007) results under stricter conditions, our results here are in line with more recent studies. For instance, Filik and Moxey (2010) used eye-tracking to measure reading times of ironic and non-ironic sentences, and subsequent text containing pronominal reference to the ironic or non-ironic text. They found faster reading times for non-ironic literal texts than for ironic ones. In addition, they found that, for ironic items, reading times of the subsequent pronominal reference indicated that both the nonsalient ironic as well as the salience-based literal interpretations were equally accessible during on-line comprehension. Such findings indicate no suppression of salience-based literal interpretations despite their incompatibility, as predicted by *the graded salience hypothesis* (Giora 1997, Giora 2003). Using strongly biasing contexts, Filik et al. (2013) also came up with similar results, while also further confirming them by using an auditory ERPs study.

Along these lines, Kowatch et al. (2013) also show that making sense of irony is quite challenging. Although ironic and literal items "were perceived as intended" (p. 307), still, literal "compliments" (*I just **love** oranges*), activating the salience-based literal interpretation, were faster to process compared to equivalent nonsalient ironic criticisms (*I just **love** apples*). Only when compared to literal criticisms (*I just **hate** apples*), did correctly interpreted ironic criticisms (which, however, amounted to only 58% of the cases), fare similarly well to these criticisms, both taking equally long to process.[11]

Findings from corpus-based studies lend further support to (the priority but even more so to) the retention of salience-based interpretations. They show that the environment of ironic utterances reflects their salience-based albeit contextually incompatible interpretations rather than their nonsalient (ironic) yet appropriate ones (Giora et al. 2014). They further testify to the difficulties in identifying

---

11 For Kowatch et al. (2013), the right comparison is between ironic and literal criticism (rather than literal compliment), since both involve similar (negative) valence.

sarcastic irony in naturally occurring discourses, despite the fact that they are cued as such (González-Ibáñez, Muresan and Wacholder 2011). And although a small set of cues, such as use of hyperbole (Kreuz and Roberts 1995) and use of interjections (e.g., *gee* or *gosh*), were shown to facilitate sarcasm **detection**, a great number of cues, such as supportive contextual information, the presence of adjectives and adverbs (but see Burgers et al. 2013), or the use of punctuation, failed to facilitate detection of sarcastic intent (Kreuz and Caucci 2007). Such findings are consistent with the view that making sense of nonsalient ironic utterances involves a misleading phase (Clark and Gerrig 1984), giving rise to salience-based interpretations initially, while further retaining them later on. They are also consistent with the relevance theoretic account (Sperber and Wilson 1986) where activating and retaining "what is said" is essential for the projection of a dissociating attitude toward it.

In sum, our results argue in favor of the temporal priority of salience-based interpretations, irrespective of contextual strength. They show that strong contextual support for nonsalient (ironic) interpretations did not sieve out accessible but inappropriate (salience-based) interpretations. Such "inappropriate" interpretations were derived initially and exclusively, regardless of strong contextual expectation for a nonsalient interpretation. They were further retained despite their inappropriateness. These results contest *the expectation hypothesis* (Gibbs 2002; Katz 2009; Campbell and Katz 2012; Pexman et al. 2000) and *the literal first model* (Grice 1975). However, they support *the graded salience hypothesis* (Giora 1997, Giora 1999, Giora 2003; Giora et al. 2007, Giora et al. 2009).

References

Akimoto, Yoritaka, Shiho Miyazawa, and Toshiaki Muramoto. 2012. Comprehension processes of verbal irony: The effects of salience, egocentric context, and allocentric theory of mind. *Metaphor and Symbol* 27(3). 217–242.
Bates, Elizabeth A. 1999. On the nature and nurture of language. In R. Levi-Montalcini, D. Baltimore, R. Dulbecco and F. Jacob (eds.). *Frontere della biologica* [Frontiers of biology – The brain of homo sapiens]. Rome: Istituto della Enciclopedia Italiana fondata da Giovanni Trecanni S.P.A., 241–265.

Bates, Elizabeth A., and Brian MacWhinney. 1989. Functionalism and the competition Model. In B. MacWhinney, and E. Bates (eds.), *The crosslinguistic study of sentence processing*. New York: Cambridge University Press, 225–256.

Brisard, Frank, Steven Frisson, and Sandra, Dominiek. 2001. Processing unfamiliar metaphors in a self-paced reading task. *Metaphor and Symbol* 16. 87–108.

Burgers, Christian, and Camiel J. Beukeboom. In press. Stereotype transmission and maintenance through interpersonal communication: The irony bias. *Communication Research*. http://dx.doi.org/10.1177/0093650214534975.

Burgers, Christian, Margot van Mulken, and Peter J. Schellens. 2013. The use of co-textual irony markers in written discourse. *Humor* 26. 45–68.

Campbell, John D., and Albert N. Katz. 2012. Are there necessary conditions for inducing a sense of sarcastic irony? *Discourse Processes* 49. 459–480.

Carpenter, Patricia A., and Meredyth Daneman. 1981. Lexical retrieval and error recovery in reading: A model based on eye fixations. *Journal of Verbal Learning and Verbal Behavior* 20. 137–160.

Clark, Herbert H., and Richard J. Gerrig. 1984. On the pretense theory of irony. *Journal of Experimental Psychology: General* 113. 121–126.

Colston, Herbert L., and Raymond W. Gibbs. 2002. Are irony and metaphor understood differently? *Metaphor and Symbol* 17. 57–60.

Dews, Shelly, and Ellen Winner. 1997. Attributing meaning to deliberately false utterances: The case of irony. In Ch. Mandell, and A. McCabe (eds.), *The problem of meaning: Behavioral and cognitive perspectives*, 377–414. Amsterdam: Elsevier.

Dews, Shelly, and Ellen Winner. 1999. Obligatory processing of the literal and nonliteral meanings of ironic utterances. *Journal of Pragmatics* 31. 1579–1599.

Duffy, Susan, A., Robin K. Morris, and Keith Rayner. 1988. Lexical ambiguity and fixations times in reading. *Journal of Memory and Language* 27. 429–446.

Filik, Ruth, and Linda M. Moxey. 2010. The on-line processing of written irony. *Cognition* 116. 421–436.

Filik, Ruth, Hartmut Leuthold, Katie Wallington, and Jemma Page. 2014. Testing theories of irony processing using eye-tracking and ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 40(3). 811–828.

Fodor, Jerry. 1983. *The modularity of mind*. Cambridge, MA: MIT Press.

Gibbs, Raymond W. 1979. Contextual effects in understanding indirect requests. *Discourse Processes* 2. 1–10.

Gibbs, Raymond W. 1986. On the psycholinguistics of sarcasm. *Journal of Experimental Psychology, General* 115. 3–15.

Gibbs, Raymond W. 1994. *The poetics of mind: Figurative thought, language, and understanding*. New York: Cambridge University Press.

Gibbs, Raymond W. 2002. A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics* 34. 457–486.

Giora, Rachel. 1997. Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics* 8. 183–206.

Giora, Rachel. 1999. On the priority of salient meanings: Studies of literal and figurative language. *Journal of Pragmatics* 31. 919–929.

Giora, Rachel. 2003. *On our Mind: Salience, Context, and Figurative language*. New York: Oxford University Press.

Giora, Rachel, Ari Drucker, Ofer Fein, and Itamar Mendelson. 2015. Default sarcastic interpretation: On the priority of nonsalient interpretations. *Discourse Processes* 52.

Giora, Rachel, and Ofer Fein. 1999a. Irony: Context and salience. *Metaphor and Symbol* 14. 241–257.

Giora, Rachel, and Ofer Fein. 1999b. On understanding familiar and less-familiar figurative language. *Journal of Pragmatics* 31. 1601–1618.

Giora, Rachel, Ofer Fein, Ann Kronrod, Idit Elnatan, Noa Shuval, and Adi Zur. 2004. Weapons of mass distraction: Optimal Innovation and pleasure ratings. *Metaphor and Symbol* 19. 115–141.

Giora, Rachel, Elad Livnat, Ofer Fein, Anat Barnea, Rakefet Zeiman, and Ido Berger. 2013. Negation generates nonliteral interpretations by default. *Metaphor and Symbol* 28. 89–115.

Giora, Rachel, Moshe Raphaely, Ofer Fein, and Elad Livnat. 2014. Resonating with contextually inappropriate interpretations in production: The case of irony. *Cognitive Linguistics* 25(3). 443–455.

Giora, Rachel, Ofer Fein, and Tamir Schwartz. 1998. Irony: Graded salience and indirect negation. *Metaphor and Symbol* 13. 83–101.

Giora, Rachel, Ofer Fein, Keren Aschkenazi, and Inbar Alkabets-Zlozover. 2007. Negation in context: A functional approach to suppression. *Discourse Processes* 43. 153–172.

Giora, Rachel, Ofer Fein, Ronie Kaufman, Dana Eisenberg, and Shani Erez. 2009. Does an "ironic situation" favor an ironic interpretation? In G. Brône and J. Vandaele (eds.) *Cognitive poetics. Goals, gains and gaps. (Applications of Cognitive Linguistics series).* Berlin/New York: Mouton de Gruyter, 383–399.

Giora, Rachel, Ofer Fein, Dafna Laadan, Joe Wolfson, Michal Zeituny, Ran Kidron, Ronie Kaufman, and Ronit Shaham. 2007. Expecting irony: Context vs. salience-based effects. *Metaphor and Symbol* 22. 119–126.

González-Ibáñez, Roberto, Smaranda Muresan, and Nina Wacholder. 2011. Identifying Sarcasm in Twitter: A Closer Look. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: short papers*, 581–586, Portland, Oregon. June 19–24, 2011.

Ivanko, Stacey L., and Penny M. Pexman. 2003. Context incongruity and irony processing. *Discourse Processes* 35. 241–279.

Jentzsch, Ines, and Werner Sommer. 2002. The effect of intentional expectancy on mental processing: a chronopsychophysiological investigation. *Acta Psychologica* 111. 265–282.

Kawamoto, Alan H. 1993. Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language* 32. 474–516.

Katz, Albert. 2009. Commentary on Does an ironic situation favor an ironic interpretation. In G. Brône and J. Vandaele (eds.) *Cognitive poetics. Goals, gains and gaps. (Applications of Cognitive Linguistics series).* Berlin/New York: Mouton de Gruyter, 401–406.

Kirby, John R. 1976. *Information processing and human abilities.* Alberta: University of Alberta.

Keysar, Boaz. 1994. The illusory transparency of intention: Linguistic perspective taking in text. *Cognitive Psychology* 26. 165–208.

Kowatch, Kristy, Juanita M. Whalen, and Penny M. Pexman. 2013. Irony comprehension in action: A new test of processing for verbal irony. *Discourse Processes* 50. 301–315.

Kreuz, Roger J., and Gina M. Caucci. 2007. Lexical influences on the perception of sarcasm. *Proceedings of the Workshop on Computational Approaches to Figurative Language.* 1–4. Rochester, NY. http://dl.acm.org/citation.cfm?id=1611529

Kreuz, Roger J., and Kristen E. Link. 2002. Asymmetries in the perception of verbal irony. *Journal of Language and Social Psychology* 21. 127–143.

Laming, Donald R. J. 1968. *Information theory of choice-reaction times*. London: Academic Press.

Laming, Donald R. J. 1969. Subjective probability in choice-reaction experiments. *Journal of Mathematical Psychology* 6. 81–120.

Loftus, Geoffrey R., and Michael E. J. Masson, Michael. 1994. Using confidence intervals in within-subject designs. *Psychonomic Bulletin and Review* 1. 476–490.

Lucariello, Joan. 1994. Situational irony: A concept of events gone awry. *Journal of Experimental Psychology* 123. 129–145.

Luka, Barbara J., and Cyma Van Petten. 2014. Gradients versus dichotomies: How strength of semantic context influences event-related potentials and lexical decision times. *Cognitive, Affective, and Behavioral Neuroscience* 14. 1086–1103.

MacWhinney, Brian. 1987. The Competition Model. In B. MacWhinney (ed.), *Mechanisms of language acquisition*, 249–308. Hillsdale, NJ: Lawrence Erlbaum.

McRae, Ken, Michael J. Spivey-Knowlton, and Michael K. Tanenhaus. 1998. Modeling thematic fit (and other constraints) within an integration competition framework. *Journal of Memory and Language* 38. 283–312.

Ortony, Andrew, Diane L. Schallert, Ralph E. Reynolds, and Stephen J. Antos. 1978. Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behavior* 17. 465–477.

Peleg, Orna, and Rachel Giora. 2011. Salient meanings: The whens and wheres. In: K. M. Jaszczolt and K. Allan (eds.) *Salience and Defaults in Utterance Processing. Mouton Series in Pragmatics*. Istvan Kecskés (gen. ed.). Berlin/New York: Mouton de Gruyter, 32–52.

Peleg, Orna, Rachel Giora, and Ofer Fein. 2001. Salience and context effects: Two are better than one. *Metaphor and Symbol* 16. 173–192.

Peleg, Orna, Rachel Giora, and Ofer Fein. 2004. Contextual strength: The whens and hows of context effects. In I. Noveck and D. Sperber (eds.), *Experimental Pragmatics*. Basingstoke: Palgrave, 172–186.

Peleg, Orna, Rachel Giora, and Ofer Fein. 2008. Resisting contextual information: You can't put a salient meaning down. *Lodz Papers in Pragmatics* 4. 13–44.

Pexman, Penny M., Todd R. Ferretti, and Albert N. Katz. 2000. Discourse factors that influence irony detection during on-line reading. *Discourse Processes* 29. 201–222.

Rayner, Keith, and Robin K. Morris. 1991. Comprehension processes in reading ambiguous sentences: Reflections from eye movements. In G. B. Simpson (ed.), *Understanding word and sentence* (Advances in Psychology, No. 77.) Amsterdam: Elsevier, North-Holland, 175–198.

Regel, Stefanie, Seana Coulson, and Thomas C. Gunter. 2010. The communicative style of a speaker can affect language comprehension: ERP evidence from the comprehension of irony. *Brain Research* 1311. 121–135.

Schwoebel, John, Shelly Dews, Ellen Winner, and Kavitha Srinivas. 2000. Obligatory processing of the literal meaning of ironic utterances: Further evidence. *Metaphor and Symbol* 15. 47–61.

Sereno, Sara C., Jeremy M. Pacht, and Keith Rayner. 1992. The effect of meaning frequency on processing lexically ambiguous words: Evidence from eye fixations. *Psychological Science* 3. 296–300.

Small, Steven L., Gary W. Cottrell, and Michael K. Tanenhaus (eds.), 1988. *Lexical ambiguity resolution and language comprehension: Computational, psycholinguistic and neurolinguistic perspectives*. San Diego: Morgan-Kaufman.

Soetens, Eric, Louis C. Boer, and Johan E. Hueting. 1985. Expectancy or automatic facilitation? Separating sequential effects in two-choice reaction time. *Journal of Experimental Psychology: Human Perception and Performance* 11. 598–616.

Sperber, Dan, and Deirdre Wilson. 1986. *Relevance: Communication and cognition*. Oxford: Blackwell.

Swinney, David A. 1979. Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior* 18. 645–660.

Tartter, Vivien C., Hilary Gomes, Boris Dubrovsky, Sophie Molholm, and Rosemarie V. Stewart. 2002. Novel metaphors appear anomalous at least momentarily: Evidence from N400. *Brain and Language* 80. 488–509.

Appendix

Items used in Giora et al. (2009)

(a) Frustrated expectation

Sagee went on a ski vacation abroad. He really likes vacations that include sport activities. A relaxed vacation in a quiet ski-resort place looked like the right thing for him. Before leaving, he made sure he had all the equipment and even took training classes on a ski simulator. But already at the beginning of the second day he lost balance, fell, and broke his shoulder. He spent the rest of the time in a local hospital ward feeling bored and missing home. When he got back home, his shoulder still in cast, he said to his fellow workers:

  "Ski vacation is recommended for your health". (Ironic)
  Everyone smiled.

(b) Realized expectation

Sagee went on a ski vacation abroad. He doesn't even like skiing. It looks dangerous to him and staying in such a cold place doesn't feel like a vacation at all. But his girlfriend wanted to go and asked him to join her. Already at the beginning of the second day he lost balance, fell, and broke his shoulder. He spent the rest of the time in a local hospital ward feeling bored and missing home. When he got back home, his shoulder still in cast, he said to his fellow workers:

  "Ski vacation is recommended for your health". (Ironic)
  Everyone smiled.

(c) No-expectation

Sagee went on a ski vacation abroad. He has never practiced ski so it was his first time. He wasn't sure whether he would be able to learn to ski and whether he will handle the weather. The minute he got there he understood it was a great thing for him. He learned how to ski in no time and enjoyed it a lot. Besides, the weather was

nice and the atmosphere relaxed. When he got back home, he said to his fellow workers:

> *"Ski vacation is recommended for your health". (Salience-based)*
> *Everyone smiled.*

Items used in Giora et al. (2007)
(d) *Barak: I finish work early today.*
   Sagit: So, do you want to go to the movies?
   Barak: I don't really feel like seeing a movie.
   Sagit: So maybe we could go dancing?
   Barak: No, at the end of the night my feet will hurt and I'll be tired.
   Sagit: ***You're a really active guy ...***
   Barak: Sorry, but I had a rough week.
   Sagit: So what are you going to do tonight?
   Barak: I think I'll stay home, read a magazine, and go to bed early.
   Sagit: ***Sounds like you are going to have a really interesting evening****.*
   Barak: So we'll talk sometime this week.

(e) *Barak: I was invited to a film and a lecture by Amos Gitai.*
   Sagit: That's fun. He is my favorite director.
   Barak: I know, I thought we'll go together.
   Sagit: Great. When is it on?
   Barak: Tomorrow. We will have to be in Metulla in the afternoon.
   Sagit: ***I see they found a place that is really close to the center.***
   Barak: I want to leave early in the morning. Do you want to come?
   Sagit: I can't, I'm studying in the morning.
   Barak: Well, I'm going anyway.
   Sagit: ***Sounds like you are going to have a really interesting evening****.*
   Barak: So we'll talk sometime this week.

Bionotes

Ofer Fein is a Senior Lecturer in Psychology in The Academic College of Tel-Aviv-Yaffo. His current research interests include psycholinguistics, psychology of sexual orientation, application of queer theory to psychology, and homophobia.

Menahem Yeari is an Assistant Professor in the School of Education at the Bar Ilan University, Israel. His primary research interests are in cognitive science, including attention, memory, and reading comprehension processes. In particular, his

research concerns the normal and abnormal cognitive processes underlying reading and discourse comprehension of typically developing and learning disabled readers, using diverse empirical methods such as behavioral measurements, eye-tracking methodology, and computational modeling.

Rachel Giora is Professor of Linguistics at Tel Aviv University. Her research areas included discourse coherence, informativeness, discourse relevance, cognitive pragmatics, language and ideology, women and language. Her recent work focuses on the psycholinguistics and neurolinguistics of figurative language (irony, sarcasm, jokes, and metaphor), context effects, optimal innovation and aesthetic pleasure, discourse negation, default nonliteral interpretation, and the notion of salience. Her book On Our Mind: Salience, Context, and Figurative Language was published by Oxford University Press in 2003. Together with Patrick Hanks she edited 6 volumes for Routledge's Critical Concepts series titled Metaphor and Figurative Language (2011).