The Defaultness Hypothesis: A quantitative corpus-based study of non/default sarcasm and literalness production

Israela Becker[*], Rachel Giora

Linguistics Department, Tel-Aviv University, Tel-Aviv 6997801, Israel

ABSTRACT

This study provides converging corpus-based evidence (originally in Hebrew) for (1) findings, supportive of the first prediction of the Defaultness Hypothesis (Giora, Givoni, & Fein, 2015b), regarding the anticipated processing speed superiority of default (spontaneous) interpretations over nondefault (context-dependent) counterparts. Indeed, as shown by Giora et al. (2015b), default negative sarcasm was processed faster than nondefault negative literalness and nondefault affirmative sarcasm, all enjoying equally strong contextual support.

Our corpus-based study here tests the second prediction of the Defaultness Hypothesis, (2) expecting relatively slow nondefaultness to be prompted by implicit and explicit cues, rejecting (as unintended) default counterparts. Findings indeed show that (i_a) speakers are sensitive to the defaultness of negative sarcasm. Hence, they implicitly preserve conceptually weak syntactic slots for nondefault interpretations; and when (i_b) wishing to be polite, speakers mute the sarcastically intended negatives by explicit metalinguistic mitigators (*to put it mildly*). Furthermore, (ii_a) when intended compositionally (often literally), speakers dismiss as unintended the default (sarcastic) interpretation of the negative constructions, by explicitly prompting a nondefault (literal) counterpart (by means of *but not Y either*). Complementarily, (ii_b) speakers intentionally fool addressees in order to humor them by using explicit cueing (*just kidding*), rejecting as unintended default (literal) interpretations of affirmative utterances while inviting nondefault (sarcastic) counterparts. Our corpus-based study, then, provides usage-based support for the Defaultness Hypothesis.

© 2018 Elsevier B.V. All rights reserved.

1. The Defaultness Hypothesis

The Defaultness Hypothesis, introduced by Giora et al. (2015b), predicts the speed superiority of processing default over nondefault (e.g., context-dependent) responses, even when constructed (rather than accessed directly from the mental lexicon). To count as a default, a response has to be automatic — activated unconditionally, initially and directly, irrespective of factors assumed to affect processing, such as degree of novelty (Giora, 1997, 2003), degree of non/literalness (Grice, 1975), degree of negation/affirmation (Horn, 1989: Ch. 3.2), or degree of strength of contextual support (Gibbs, 1994). Additionally, when tested for degree of defaultness (e.g., by means of measuring processing speed), items have to be free of cues inviting non/literalness, so that ambiguity is allowed. They should therefore be (i) novel/noncoded, (ii) free of semantic anomaly or

* Corresponding author.
  E-mail addresses: israela.becker@gmail.com (I. Becker), rachel.giora@gmail.com (R. Giora).

internal incongruity, and (iii) free of contextual support, thereby rendering a preference optional. According to the Defaultness Hypothesis, then, the speed superiority of a response over an alternative counterpart is affected by defaultness, and nothing but defaultness.

In Giora et al. (2015b), degree of defaultness was established when items (meeting conditions i—iii above) were presented in isolation and rated for degree of interpretive preference (Experiment 1.1) and sarcasm (Experiment 1.2). Findings resulted in

(a) 2 default, automatic interpretations — negative sarcasm (*s/he is not the most intelligent person in the room*, meaning 's/he is dumb') and affirmative literalness (*s/he is the most intelligent person in the room*, meaning 's/he is the smartest'); and

(b) 2 nondefault (e.g., context-dependent) interpretations — negative literalness (*s/he is not the most intelligent person in the room*, meaning 'others are more intelligent than her/him') and affirmative sarcasm (*s/he is the most intelligent person in the room*, meaning 's/he is dumb').[1]

Hence, when presented in equally strong contexts, supportive of their respective interpretations, both **default** interpretations (**negative sarcasm** and **affirmative literalness**; see Table 1) were processed faster than their **nondefault** counterparts (**negative literalness** and **affirmative sarcasm**; see Table 1). Specifically, **default negative sarcasm** was processed faster than its nondefault alternative counterparts — **nondefault affirmative sarcasm** and **nondefault negative literalness**. Similarly, **default affirmative literalness** was processed faster than its nondefault alternative counterparts — **nondefault affirmative sarcasm** and **nondefault negative literalness** — all equally strongly supported by contextual information (see Giora et al., 2015b: Experiment 2. For corroborating evidence, using eye tracking during reading English items, see Filik et al., 2018; for further corroborating evidence, using a divided visual field paradigm, involving a lexical decision task in Hebrew, see Giora et al., 2018a).

Table 1
The default and nondefault interpretations of affirmative and negative utterances.

	Default interpretation	Nondefault interpretation
Negative utterance	Sarcastic	Literal
Affirmative utterance	Literal	Sarcastic

Based on the (predicted and verified) processing speed superiority of defaultness over nondefaultness, the Defaultness Hypothesis *further* predicts the involvement of cueing in prompting nondefaultness, while rejecting defaultness.[2] Nondefault responses, then, are expected to rely heavily on cueing; however, intensifying or inviting default (automatic) responses will be rare, as will be cues rejecting nondefaultness (the latter involving a slow, non-automatic response).[3]

Given the (relative) slow speed of deriving nondefault interpretations (see Giora et al., 2015b), the present study (using natural Hebrew discourse) will test the second prediction of the Defaultness Hypothesis, expecting nondefault rather than default interpretations to be explicitly prompted by cueing.

1.1. Cueing inviting nondefaultness

Note that, so far, Giora and colleagues have focused primarily on experimentally testing the first prediction of the Defaultness Hypothesis, related to the speed superiority of default interpretation over nondefault counterparts. For instance, Giora and colleagues measured the processing speed of some negative constructions and their affirmative counterparts (originally in Hebrew), such as *X s/he is not* and *X s/he is yes*[4] (Giora et al., 2013: Experiment 3), *X is/is not her/his forte/prominent strength/dominant attribute* (Giora et al., 2015a: Experiments 2 and 4), and *s/he is/is not particularly/the most X* (Giora et al., 2015b: Experiment 2). In these constructions, the *X* slot accommodates either an adjective-phrase or a noun-phrase, each of which hosts a positively-oriented concept. (On defining such constructions as strongly attenuating highly *positive* concepts, see Giora et al., 2018b). Before running the online processing experiments, Giora and colleagues used offline rating tasks, showing that the default preferred interpretation of these negative constructions—*X s/he is not* (Giora et al., 2013: Experiment 2), *X is not her/his forte/prominent strength/dominant attribute* (Giora et al., 2015a: Experiments 1 and 3), and *s/he is not the most X* (Giora et al., 2015b: Experiment 1)—is sarcastic; their nondefault interpretation is literal.

---

[1] *Sarcasm* refers here to what is termed in the literature as verbal irony (see, e.g., Gibbs, 1994; Sperber, 1984).

[2] We use *cue* as an umbrella term covering discourse markers as well as all kinds of "salient discourse patterns" (see Ariel, 2008: Ch. 5), whether grammaticized or not.

[3] Note that here we focus on nondefault, noncoded, constructed *interpretations*. Hence, cueing, inviting nondefault coded *meanings* low on salience rather than default counterparts high on salience, is not discussed here, as both are accessed directly from the mental lexicon rather than formed on the fly (see, e. g., Giora, 1997, 2003; however, on cueing nondefault meanings, see Givoni et al., 2013).

[4] In Hebrew, the affirmative in such cases is explicit.

According to prevailing views (e.g., Grice, 1975; Horn, 1989: Ch. 3.2; and references therein), the two factors, negation and figurativeness (e.g., negative sarcasm, negative metaphor), are expected to slow down the processing speed of these constructions compared to their nondefault counterparts, whether affirmative or literal.

In contrast to those prevailing views, Giora and colleagues consistently showed that these two factors need not always slow down processing. The processing speed of the sarcastically interpreted negative constructions, in *all* these studies, is faster than that of their nondefault counterparts − whether nondefault affirmative sarcasm or nondefault negative literalness (see Giora et al., 2015b). Similarly, the processing speed of default metaphorically interpreted negative constructions is faster than that of literally interpreted negative counterparts (see Giora et al., 2013). When it comes to processing, then, it is only defaultness that counts.

Ultimately, the results presented by Giora and colleagues provide strong support for the Defaultness Hypothesis. In fact, the Defaultness Hypothesis is the only theory that can account for *all* these results taken together.

Crucially, the Defaultness Hypothesis has been tested (and supported) mostly by online (but also by offline) experiments, focused on *comprehension*. As we believe that "[l]anguage use is characteristically a joint activity" of producers and comprehenders (Wasow, 1997: 360; see also Du Bois, 2003, 2014), the current study looks into natural language *production*, seeking to further test the Defaultness Hypothesis by adopting a methodologically different angle − a quantitative corpus-based approach, given that corpora, rather than elicitation experiments, are the best manifestation of production.

In order to test the Defaultness Hypothesis in terms of natural use, we examine the second prediction (of the Defaultness Hypothesis) suggesting that cueing *nondefault* negative literalness will be more prevalent than cueing *default* negative sarcasm; and cueing *nondefault* affirmative sarcasm will be more prevalent than cueing *default* affirmative literalness (see (c−d) below). However, before testing this prediction, we provide corroborating, usage-based evidence supportive of the defaultness of negative sarcasm (see (a−b) below).

Our study, then, involves 4 pieces of corpus data in Hebrew, testifying to speakers' sensitivity to defaultness, made noticeable when (a) they implicitly refrain from embedding them in weak syntactic positions; or (b) when they preface the utterances by a mitigating cue, thereby acknowledging the bluntness of the default interpretation; or (c−d) when they reject the default unintended interpretation in favor of explicitly prompted nondefault counterpart. Specifically, we show that

a) speakers refrain from embedding certain negative constructions in linguistic environments, which host a weak message, thus implicitly acknowledging the strong negative impact of their default sarcastic interpretation;
b) speakers tone down the strong blunt message of the default sarcastic interpretation of these negative constructions by using a metalinguistic mitigator (e.g., *to put it mildly*) rather than an intensifier (e.g., *not to be blunt, but*), thus signaling that they are aware of the harsh message these constructions convey;
c) speakers prompt the nondefault literal interpretation of negative constructions (*s/he is not the kindest person around*) by rejecting their default unintended sarcastic interpretation, using explicit cues, such as *but not Y either* (as in *but not inhumane either*);
d) speakers prompt the nondefault sarcastic interpretation of (mostly) affirmative utterances by explicitly rejecting their default compositional (here, literal) interpretation, using cues such as *stǽːm* (Hebrew for 'not seriously') or *just kidding*, as in "… *the plan is … to expose the banking industry in Israel to competition … stǽːm!, just kidding!* …".[5,6]

This paper is therefore organized in the following way: As most of this study is devoted to examining some specific negative constructions, previously shown to be interpreted sarcastically by default, we start by listing these constructions (henceforth, the negative constructions studied here; see Section 2). We then present, in detail, the considerations taken when choosing a specific corpus for our study (see Section 3). Next, we present 4 pieces of corpus-based evidence attesting to speakers' sensitivity to the default interpretation of utterances: The first two pieces of evidence attest to the defaultness of the sarcastic interpretation of the negative constructions studied here (see a−b above), conveyed by both implicit (see Section 4) and explicit (see Section 5) cues. The next two pieces of evidence discuss explicit cues speakers employ, which invite a nondefault literal interpretation of negative constructions, while rejecting (as unintended) their default sarcastic interpretations (see c above), and which further invite a nondefault sarcastic interpretation of affirmative items, while rejecting (as unintended) their default literal counterparts (see d above). Such uses further disclose speakers' sensitivity to the defaultness of certain interpretations (see Section 6). Finally we wrap up (see Section 7).

2. Default negative sarcasm

As mentioned above, Giora et al. (2013, 2015a,b) studied several negative constructions modifying highly positive concepts, using online (and offline) experiments. Results show that these specific constructions are interpreted sarcastically by default:

---

[5] All the examples given throughout this paper were extracted from *HeTenTen*, a web-corpus of Modern Hebrew, described in detail in Appendix A.
[6] On prompting nondefault affirmative sarcasm by cueing it, see e.g., Attardo (2000); Booth (1974); Filik et al. (2016); Haiman (1998); Kreuz (1996); Muecke (1978); Sulis et al. (2016); Thompson and Filik (2016); Veale (2018). On cueing novel nondefault nonliteralness, see e.g., Katz and Ferretti (2003).

a) *X s/he is not* (*alert s/he is not*; see Giora et al., 2013);
b) *X is not her/his forte/prominent strength/dominant attribute/strong suit* (*alertness is not her/his forte/prominent strength/dominant attribute/strong suit*; see Giora et al., 2015a);
c) *s/he is not the most X* (*s/he is not the most alert person around*; see Giora et al., 2015b).

Note that the constructions in (a) and (b) are structurally marked, while that in (c) is structurally unmarked. Since a quantitative corpus-based study requires a considerable number of instances, we focus here on this specific unmarked construction (c), *s/he is not the most X* and its variants — *s/he is not particularly X, s/he is not extremely X*, and *s/he in not really X*. These constructions are expected to be prevalent in a corpus and provide many more instances than the marked alternatives — *X s/he is not* and *X is not her/his forte/prominent strength/dominant attribute/strong suit* constructions.

Additional constructions fitting into this same natural category (of unmarked constructions) are *s/he is not so X, s/he is not exactly X, s/he is not entirely X*, and *s/he is not too X*. These additional constructions, as well as some structurally idiosyncratic examples such as *they are not candidates for winning the Noble prize* (which are expected to behave in a similar manner to the other negative constructions studied here), are considered only in Section 5.

But are the relatively prevalent unmarked negative constructions discussed above *also* absolutely prevalent? The answer to this question determines which corpus can provide us with a statistically fair amount of relevant instances adequate for a quantitative corpus-based study.

3. The corpus

Tannen (1984: Ch. 6) and Gibbs (2000) found that irony constitutes 7% or 8% of all conversational turns among friends (Americans in their 20s and 30s). Gibbs further investigated five main forms of irony, of which the understatement category (to which the negative constructions studied here belong) is the least frequent one — ~0.2%. Both Tannen and Gibbs considered *all* kinds of understatements, regardless of their surface form. So if understatements are generally rare with respect to other forms of irony, let alone literally interpreted utterances, then all the more so with regard to specific understated constructions such as the negative constructions studied here.

In order to assess the size of the corpus needed for compiling a statistically reasonable number of the negative constructions studied here, we closely inspected the Corpus of Spoken Israeli Hebrew.[7] Not only is it a corpus of conversations among friends or peers like Tannen's corpus, but it is also of a comparable size: Tannen's corpus comprises of 2.7×10^4 tokens and the Corpus of Spoken Israeli Hebrew comprises of 4×10^4 tokens. The number of the negative constructions studied here, found in the Corpus of Spoken Israeli Hebrew, is extremely scarce — altogether 2 (see Table 2).

Table 2
The (lack of) prevalence of the negative constructions studied here in the Corpus of Spoken Israeli Hebrew.

The construction	# of instances
s/he is not the most X	1
s/he is not particularly X	0
s/he is not extremely X	0
s/he is not really X	1

If, then, a corpus of around 4×10^4 tokens (i.e., the Corpus of Spoken Israeli Hebrew) provides us with *only* 2 instances of the relevant negative constructions studied here, whereas hundreds (or even thousands) of such constructions are needed in order to achieve reliable statistical results, then a corpus of several million tokens should be considered. This estimate is in line with Hary and Izre'el (2003), who argued that "[a] corpus of five million spoken words seems to be of sufficient size to represent well both the overall structure and the **specific features** of most linguistic varieties of potential interest" (p. 196; emphasis added). But even a corpus of several million tokens might not do, as we are interested in specific negative constructions accompanied by very specific discourse cues, such that constitute tools to test the Defaultness Hypothesis. These negative constructions, accompanied by specific discourse cues, are constrained sequences. As such, they are rare and call for larger corpora than a corpus of several million tokens. *HeTenTen*, a web-corpus of Modern Hebrew, is such a large corpus, larger than the Corpus of Spoken Israeli Hebrew (and in fact of any other corpus of Modern Hebrew), and comprises about 1×10^9 tokens. Hence, it is expected to contain a reasonable number of the negative constructions we are after here, accompanied by the very specific discourse cues which are of interest to us. HeTenTen is, therefore, the corpus used throughout this study. (For a detailed description of HeTenTen, see Appendix A).

Having selected an adequate corpus, we start by presenting corpus data testing the Defaultness Hypothesis.

---

[7] http://cosih.com/.

4. Implicitly signaling defaultness by means of avoiding weak syntactic slots

As mentioned earlier, we aim to test the Defaultness Hypothesis in usage-based terms, while providing converging corpus-based evidence to online findings. Recall, that the latter attested to the defaultness of the sarcastic interpretation of the negative constructions studied here by means of processing speed (and offline experiments).

We, therefore, investigate whether, as predicted by the Defaultness Hypothesis, the (default) interpretation of these specific negative constructions constitutes a strong enough (i.e., sarcastic) proposition, assuming an extreme end of a conceptual scale. This inspection is performed by examining a specific syntactic environment, *W in fact S* (*W*eak in fact *S*trong), as in *s/he is good, in fact, s/he is excellent.* Given this *W in fact S* syntactic environment, the results of this examination will enable us to decide whether the negative constructions studied here, hosting a positive concept, are interpreted sarcastically (i.e., strongly) by default.

A default sarcastic interpretation − an end-of-the-scale interpretation, such as 's/he is dumb' derived from *s/he is not really smart* − may be implicitly highlighted by the absence of such negative constructions from the *W* (*Weak*) slot of *W in fact S* (*Strong*) syntactic environment.

Let us first elaborate on the *W in fact S* syntactic environment.

4.1. The W in fact S syntactic environment

The connective *in fact* (as well as e.g., *indeed*, *actually*, *as a matter of fact*, and *what's more*), in the *W in fact S* syntactic environment, signals that the utterance that precedes it (i.e., *W*) is conceptually weaker than the utterance that follows it (i.e., *S;* see Horn, 1989: ff. 231). Horn noticed that syntactic environments such as "P_i, *indeed/in fact/and what's more* P_j" (p. 234, example 50(b)) would easily host scalar predicates[8] such as <good, excellent> to produce a coherent utterance such as example (1):

(1) ✓ He is good, in fact, he is excellent

But the result of a reversed order of the scalar predicates, that is, P_j *in fact* P_i, will be unacceptable, as in example (2):

(2) # He is excellent, in fact, he is good

In terms of a position on a conceptual scale, the *W* slot can host utterances which are conceptually weaker than utterances occupying the *S* slot. Consequently, the *W* slot cannot host an utterance that occupies the uttermost end of the scale, namely, a strong one (see Schwenter and Traugott, 2000; Traugott and Dasher, 2002: Ch. 4; Oh, 2000; and Aijmer, 2013: Ch. 3 who argue that the *W* slot in the *W in fact S* syntactic environment can host an utterance which is also *rhetorically*—and not just conceptually—weaker than the utterance that occupies the *S* slot).

Horn (1989) contended that the specific use of the utterance, occupying the *W* slot, Q-implicates ("Say as much as you can, and imply no more") that the speaker could not commit to any stronger version of this utterance. The subsequent *in fact* and its like promptly induce the cancellation of this scalar Q-implicature (of the utterance occupying the *W* slot), suggesting that the utterance occupying the *S* slot is a stronger version of the utterance occupying the *W* slot.

As for Hebrew — Bardenstein (2019) shows that the Hebrew *be-ʃecem* and *le-maʃase* are equivalent to *in fact*, as they serve as mediators between two consecutive utterances, of which the former is a weaker version of the latter.

4.2. Which slot of the W in fact S syntactic environment will the negative constructions studied here occupy?

Assume that the strong *S* slot in the *W in fact S* syntactic environment is already occupied, i.e., blocked. In such a case, the only slot, open to hosting any of the negative constructions studied here, is the weak *W* slot. Under such circumstances, a nondefault, literally intended negative construction, occupying a midpoint on a scale, might fit into this open weak *W* slot. In contrast, a default sarcastically intended negative construction, assuming the utmost endpoint of a scale, will be too strong to occupy this weak *W* slot.

Before testing this prediction, one should, however, note that absence from the weak *W* slot of any of the negative constructions studied here could *also* be the consequence of the small number of occurrences of these negative constructions, rather than the impact of their sarcastic, end-of-the-scale interpretation. To avoid such a state of affairs, we focused on the *s/he is not really X* variant of the negative constructions studied here, since it is the most prevalent construction of its kind in the corpus (see Table 3, in which the *X* slot of the negative constructions, hosting an adjectival phrase, reflects the prevalence of these constructions in the corpus, regardless of whether they host an adjectival phrase, a noun phrase, or a verb phrase).

---

[8] Horn defined scalar scales by entailment: "P_j outranks P_i on a given scale iff a statement containing an instance of the former unilaterally entails the corresponding statement containing the latter" (p. 231).

Hence, if no instance of *s/he is not really X*, hosting a positive concept and occupying the weak *W* slot of *W in fact S* syntactic environment is found, then this could be accounted for *only* by its default strong sarcastic end-of-the-scale interpretation, rather than its low frequency.

Table 3
The frequency in HeTenTen of the negative constructions studied here hosting an adjectival phrase in the *X* slot.

The negative construction	# of instances
s/he is not really X	14,159
s/he is not the most X	4000
s/he is not extremely X	761
s/he is not particularly X	96

The next step, then, is to examine whether instances of the *W in fact S* syntactic environment, in which the strong *S* slot is blocked, host in the weak *W* slot the negative constructions studied here.

4.3. The subset of W in fact S examples in which the S slot is blocked

In order to focus on instances of *W in fact S*, in which the strong *S* slot is blocked, we decided to test instances in which the extreme value sequence *s/he is **really not** X*, in which the intensifier scopes over the negator (rather than a negator scoping over an intensifier, as in the negative constructions studied here, *s/he is **not really** X*), occupies (thus, blocks) the strong *S* slot, exemplified in (3) and (4) (both originally in Hebrew):

(3) Is he looking for a creative publicity system? Not sure. *Not really*. In fact, *certainly* [*really*] *not.*
(4) After all we are *not really* in Downtown Manhattan, viz. we are *really not.*

This constraint (*s/he is **really not** X*, occupying the strong *S* slot) resulted in 24 cases of the form *s/he is **not really** X, in fact, s/he is **really not** X*, in which the *s/he is **not really** X* occupies the weak *W* slot (two of which are (3) and (4) above).

These findings seem to testify to the conceptual weakness of *s/he is **not really** X*, that is, to a default literal interpretation of *s/he is **not really** X*. These findings also seem to be in conflict with prior experimental results showing that the default interpretation of the negative constructions studied here is sarcastic. But is it indeed the case?

It has been previously shown that when negative constructions host *non*-positive concepts, they are interpreted as literal or as weaker ironies, as in *s/he is not really stupid* (see e.g., Dews and Winner, 1995; Goldenberg, 2011; Schwoebel et al., 2000). As such, they occupy some midpoint on a conceptual scale (rather than an opposite endpoint of that scale), constituting a mitigated proposition, and therefore not precluded from occupying the open weak *W* slot.

But what about the negative constructions studied here hosting positive concepts? If literally interpreted, then similar to the negative constructions hosting non-positive concepts, they are expected to occupy some midpoint on a scale, and therefore should not be precluded from occupying the open weak *W* slot, either. However, if they are interpreted sarcastically, as shown previously by Giora and colleagues (Giora et al., 2013, 2015a,b), then they are expected to be displayed at the endpoint of this scale. As such, they are barred from occupying the weak *W* slot, which hosts only weaker mitigated propositions. These predictions are summarized in Table 4, using *s/he is not really X* as representing the negative constructions studied here.

Table 4
The expected frequency of the *W in fact S* syntactic environment, hosting any of the negative constructions studied here in the weak *W* slot, as a function of their interpretation which, in turn, is a function of the polarity of the concept they host.

Polarity of the concept occupying the X slot in *s/he not really X*	Interpretation of *s/he not really X*	How many *'s/he is not really X, in fact S'* are expected to be found?
Positive	Sarcastic	None
	Literal	Some
Non-positive	Literal	Some

Accordingly, if we find that the 24 cases of *s/he is **not really** X*, occupying the weak *W* slot, host positive concepts, then they make a mitigated proposition just as *s/he is **not really** X*, where *X* is a *non*-positive concept. In other words, *s/he is **not really** X* is interpreted as literal across-the-board, irrespective of the polarity of the concept in its scope. However, if we find that these 24 cases of *s/he is **not really** X*, occupying the weak *W* slot, do not host positive concepts, then there are practically no cases of *s/he is **not really** X* occupying the weak *W* slot, while hosting positive concepts. This, then, could be attributed only to their default, strong end-of-the-scale sarcastic interpretation.

To determine whether the 24 concepts, occupying the *X* slot of *s/he is **not really** X* (in *s/he is **not really** X, in fact, s/he is **really not** X*), which, in turn, occupies the weak *W* slot of *W in fact S* syntactic environment, are positively oriented or not, we conducted an offline web-experiment. In this experiment, participants were asked to compare the polarity of these 24 concepts to the polarity of 24 alternative concepts, occupying the *X* slot in sarcastically interpreted instances of *s/he is **not really** X* (used for a completely different task), in which the 24 alternative concepts were previously rated (by 3 expert judges) as positive.

4.4. Method

4.4.1. Participants

Forty-four volunteers, native speakers of Hebrew, (17 women and 27 men), aged 24–78 ($M = 40.6$, $SD = 13.5$), were recruited via the web. They were all highly educated (16–28 years of education altogether) with expertise in the following disciplines: Eighteen participants graduated from faculties of Exact Sciences or Engineering, 6 were linguists, 6 graduated from faculties of the Humanities but did not specialize in linguistics, another 2 graduated from faculties of Social Sciences (Psychology not included), 1 from the faculty of Natural Sciences, 1 from the faculty of Management, and 1 from the faculty of Medicine. The remaining 9 participants did not specify their field of expertise.

4.4.2. Materials

The participants were presented with 62 items: The first 4 items served as buffers; the following 58 items consisted of the 24 concepts that filled in the *X* slot in the 24 *s/he is **not really** X in fact s/he is **really not** X* utterances, extracted from HeTenTen (henceforth, the experimental items); 24 more items included concepts which occupied the *X* slot in 24 pseudo-randomly selected instances from an unrelated list of 53 instances of *s/he is **not really** X* (see Section 5), judged as sarcastic by 3 expert judges versed in the field of sarcasm (henceforth, the control items). These 3 expert judges also judged the concepts that these constructions host as positively biased; 10 more items served as fillers.

All the items were pseudo-randomly ordered (including the ordering within the set of the 4 buffer items) so that each participant had a list of items uniquely arranged just for her/him.

4.4.3. Procedure

Participants received a web-link to the experiment. The first two screens of the experiment displayed the instructions, including two examples. The participants were asked to rate any item as either positive, neutral, or negative. Then each participant saw 62 screens. Each screen contained a single word or phrase (in italics), centered at the top of the screen and followed by a nominal three button scale (see Fig. 1 below). The participants were allowed unlimited time to reflect upon their ratings. Note that participants were presented with nominal categories, — negative, neutral, and positive — concealing an ordinal scale of −1, 0, +1, respectively, the latter used for the following data analysis.
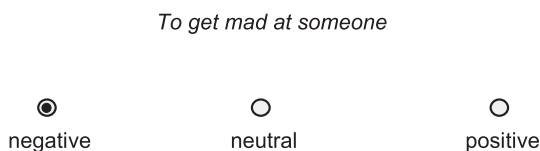
To get mad at someone

◉	○	○
negative	neutral	positive

Fig. 1. An exemplar screen of the web-experiment in which a word or a phrase, in italics, could be rated as positive, neutral, or negative.

4.5. Results and discussion

Results show that the polarity of the 24 experimental items (i.e., those filling in the weak *W* slot of *s/he is **not really** X in fact s/he is **really not** X* utterances) was rated as lower in positivity than that of the 24 control items. Recall that, the control items, taken from an unrelated task, were previously rated by expert judges as involving a positive concept (hosted in the *X* slot of sarcastic *s/he is **not really** X* utterances). Results were significant by both, participant (*t*) and item (*U*) analyses. Participant analysis reveals that the mean score of the experimental items was low, indicating neutrality ($M = 0.1506$, $SE = 0.0205$), compared to the mean score of the control items, indicating positivity ($M = 0.6581$, $SE = 0.0231$). This difference, 0.50758, BCa 95% CI [0.45845, 0.55670], is significant $t(43) = 20.839$, $p < 0.001$, and represents an extremely large size effect (Cohen's $d = 3.15$). The item analysis reveals that the median of the experimental items was low, indicating neutrality ($Mdn = 0.25$), compared to the median of the control items, indicating positivity ($Mdn = 0.7727$), $U = 105$, $z = 3.777$, $p < 0.001$, $r = 0.545$.

Such results clarify the conditions under which *s/he is **not really** X* constructions can occupy the weak *W* slot of the *W in fact S* environment. As these 24 instances of the *s/he is **not really** X* construction host a *non*-positive concept in the *X* slot, they make a mitigated expression, i.e., such that is literally interpreted, and therefore can occupy the weak *W* slot of the *W in fact S* environment, without involving any conflict between the intensity of the item (which is mitigated) and the intensity of the hosting slot (which can tolerate mitigated propositions only). These findings attest to the defaultness of the sarcastic interpretation of the negative constructions studied here, hosting positive concepts. Given that they are highly intense, and, as such, in conflict with the kind of intensity of the weak *W* slot of the *W in fact S* syntactic environment, their absence from

this weak slot (which can tolerate utterances of a conceptually mitigated nature *only*), testifies to the robustness of their default sarcastic interpretation.

If these results are generalized to any of the other negative constructions studied here, then we can conclude that, when hosting positive concepts, these constructions will all be interpreted sarcastically by default (as attested to experimentally by Giora et al., 2013, 2015a,b, 2018a,b; Giora, submitted; Giora et al., submitted; Giora et al., 2017).

5. Explicitly signaling defaultness by using a metalinguistic comment 'to put it mildly' (Hebrew *lomar be-ʃadinut)*

Following the *implicit* signaling of the default sarcastic interpretation of the negative constructions studied here (presented in the previous section), we further introduce a metalinguistic device, intended to *explicitly* cue this interpretation in an attempt to mitigate the blunt, critical message conveyed by the negative constructions studied here.

According to politeness theories, speakers wish to be polite as a matter of course (e.g., Brown and Levinson, 1987). However, occasionally, they need to convey a harsh message. If a message bites, then given the attempt to be polite and save face, speakers would try to tone it down. Is this indeed the case with the negative constructions studied here? Do they tend to be prefaced by means of (variants of) the colloquial metalinguistic comment *to put it mildly*, as a means to attenuate a harsh message, or opt, instead, for an alternative — *(not) to be (too) blunt/rude, but* — making explicit the negativity of the default sarcastic interpretation?

According to the Defaultness Hypotheses, if the negative constructions studied here are indeed sarcastic by default, even if expressed in an understatement manner, then, in line with politeness theories, they might call for a mitigating device, toning down their harsh message. In contrast, defaultness needs no amplification. Hence, cues highlighting the negativity of the specific sarcastic items should be rare; given the defaultness of the negative constructions studied here, there is actually no need to further spell it out.

In order to test these predictions, following from the Defaultness Hypothesis and politeness theories, we extracted from HeTenTen 171 instances of the negative constructions studied here (see example 5), as well as similar variants, such as, *s/he is not so X, s/he is not exactly X, s/he is not entirely X*, and *s/he is not too X* (see example 6), syntactically marked constructions of the type *X s/he is not* (see example 7), and even idiosyncratic negative constructions (see example 8), all preceded or followed by *to put it mildly* or any of its variants. (For a description of the extraction procedure, see Appendix B).[9]

(5) Most of us lead a way of life which … *how to put it mildly*, **cannot be considered particularly healthy!**
(6) A simple calculation … shows that each employee sheds 200 g of his weight, which is, *how to put it mildly*, **not exactly something to write home about**, and certainly not something to write in the newspaper about …
(7) *How to put it mildly*, **a superb writer she is not**.
(8) *Let's put it mildly*, even after the fake documents **he doesn't come out as the Righteous among the Nations**. At best, he was not found guilty of being an active collaborator of Himmler.

At the same time, we managed to extract from HeTenTen nothing but a single instance of the negative constructions studied here (including the extensive list of marked ones and idiosyncratic instances as in examples 7—8 above), following *I don't want to be too blunt, but* (one of several *(not) to be (too) blunt/rude, but* variants) as in example (9) (For a description of the extraction procedure, see Appendix B).

(9) *I don't want to be too blunt, but*[10] **this conduct is not the most courageous thing I have come across**.

To make sure that speakers consider the negative constructions studied here as sarcastic when preceded (or followed) by *to put it mildly* or by *(not) to be (too) blunt/rude, but*, we asked 3 judges, versed in the field of sarcasm, to classify each of the 172 instances of the negative constructions, (including 171 items juxtaposed to *to put it mildly* and 1 item following *I don't want to be too blunt, but*), as either sarcastic or literal. Most of the 160 items, following *to put it mildly*, specifically 73.1% (125/171; binomial test, $p < 0.001$), were judged as sarcastic by all 3 judges. *All* 171 items, following *to put it mildly*, were judged as sarcastic by at least 2 judges. The single item following *I don't want to be too blunt, but* (see example 9 above) was judged as sarcastic by all 3 judges.

Taking into account the prevalence of *to put it mildly* (and its variants) and the prevalence of *(not) to be (too) blunt/rude, but* (and its variants) in HeTenTen (for a description of the procedure by which we determined the entire population of each cue in

---

[9] There is a similar metalinguistic cue — *and that's a (gross) understatement* — serving the same pragmatic function as the *to put it mildly* cue. Although the former is more frequent than the latter, we focused on the latter rather than on the former. Recall, that we are interested in metalinguistic cues such as those two, preceding or following the specific kind of negative constructions studied here (i.e., *s/he is/is not particularly/the most/very/really X* and the like). Our preliminary scan for instances of any of these two linguistic cues preceding or following utterances in which highly positive concepts are attenuated, produced equal amounts of these two cues preceding or following such utterances (~20% of all items). This means that both, *and that's a (gross) understatement* and *to put it mildly*, are associated with such utterances. However, whereas 70% of the utterances preceded or followed by *to put it mildly* are of the kind of negative constructions studied here (i.e., *s/he is/is not particularly/the most/very/really X* and the like), only 20% of the utterances preceded or followed by *and that's a (gross) understatement* are of that kind. That is why we focused on *to put it mildly*.

[10] *Not to be (too) blunt/rude, but* is always followed by a blunt utterance despite the seemingly mitigated nature of this remark (see Ran, 2015).

the corpus, see Appendix B), there are significantly more instances of the negative constructions studied here, preceded (and occasionally followed) by *to put it mildly* (and its variants, 171/872 = 19.6%) than instances of these negative constructions preceded by *(not) to be (too) blunt/rude, but* (and its variants, 1/78 = 1.3%), χ^2(for homogeneity) = 16.22, $p < 0.001$. These results testify to the defaultness of the sarcastic interpretation of the negative constructions studied here which may benefit from additional mitigation (rather than additional intensification) to compensate for the bluntness of the message conveyed, in order to comply with accepted codes of human conduct, viz. politeness.

So far, we have presented corpus-based evidence highlighting (both implicitly and explicitly) the default sarcastic interpretation of the negative constructions studied here. In the next section we present 2 explicit cues aiming at prompting nondefault interpretations by rejecting unintended default alternatives — negative sarcasm and affirmative literalness (see Giora et al., 2015b).

6. Explicitly rejecting unintended default interpretations while prompting nondefault counterparts

6.1. 'But not Y either' (Hebrew aval gam lo): rejecting default sarcastic interpretations of the negative constructions studied here while inviting nondefault literal counterparts

Given that the negative constructions studied here are interpreted sarcastically by default, then, according to the Defaultness Hypothesis, an addressee, encountering *s/he is not the most exciting person in the world*, will instantly interpret it sarcastically as 'she is an extremely boring person', as shown by Giora et al. (2015b) and illustrated in Fig. 2.
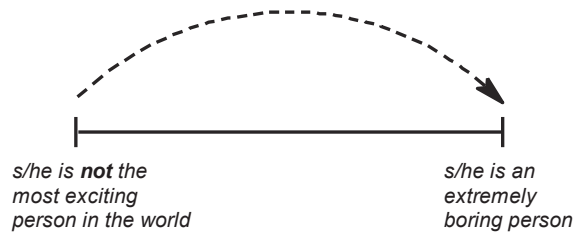

Fig. 2. As indicated by the dashed arrow, the default interpretation of *s/he is not the most exciting person in the world* occupies the extreme opposite end of the scale, conveying the sarcastic interpretation ('s/he is an extremely boring person').

In other words, sarcastically interpreted stimuli, such as *s/he is not the most exciting person in the world*, communicate the opposite of what is said *by default*. Hence, *s/he is not the most exciting person in the world* occupies the extreme end of a polarity scale rather than an arbitrary mid-point on that scale. Still, if a speaker's intent is to convey the nondefault literal message of the negative construction, then she is likely to explicitly signal to the addressee that her message should be taken compositionally (e.g., literally) rather than sarcastically. To that end, the speaker is likely to use a cue, following the negative construction, aiming at rejecting the default (unintended) sarcastic interpretation while inviting the nondefault, compositional (here literal) alternative. One such cue is *but not Y either* — a contrastive salient discourse pattern (Ariel, 2008: Ch. 5) — which is likely to reject the default sarcastic interpretation of the utterance, while inviting its nondefault counterpart (as illustrated in Fig. 3).[11]
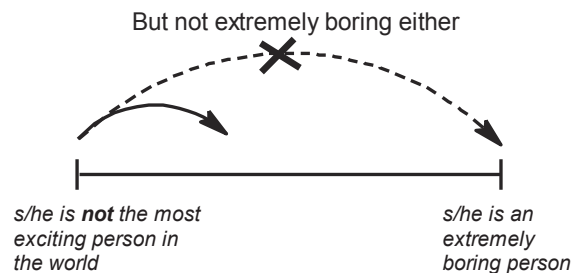

Fig. 3. The default interpretation of *s/he is not the most exciting person in the world* (indicated by the dashed arrow) is rejected by a cue (indicated by the cross over the dashed arrow), inviting a nondefault intermediate position on the polarity scale, conveying a literal message (as indicated by the solid arrow), thus suggesting that 's/he is an exciting person but not the most exciting one'.

---

[11] Note, that the *Y* slot in *but not Y either* accommodates an antonym or (equally so) an ad-hoc antonym of the concept accommodated by the *X* slot of the negative construction.

At the same time, if the negative constructions studied here are interpreted sarcastically by default, then according to the Defaultness Hypothesis, they are not expected to be followed by any cue that would steer their interpretation towards the extreme opposite end of the scale. After all, this opposite is already their *default* interpretation. These negative constructions are, therefore, not expected to be followed by absolute negators, such as *not even close/not even getting close/even quite (fairly) remote/even the opposite* or *if at all*, since such sequences are informationally redundant. Items involving such absolute negators are illustrated in Fig. 4 and are not expected to be prevalent.
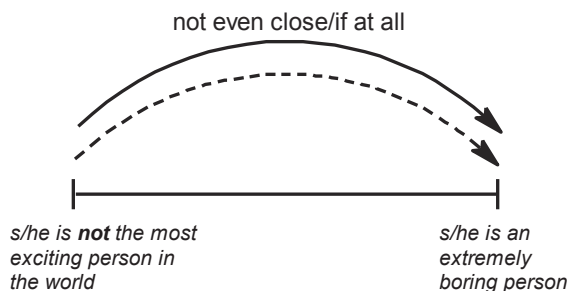

Fig. 4. *S/he is not the most exciting person in the world* is interpreted as conveying the extreme opposite of what is said, meaning 's/he is an extremely boring person' (as indicated by the dashed arrow). The additional *not even close/if at all* (indicated by the solid arrow) is superfluous here, as it amplifies an already amplified value.

Do speakers, then, reject (by means of *but not Y either*) the default sarcastic interpretation of the negative constructions studied here when they wish to use these constructions literally, but hardly ever intensify this default sarcastic interpretation (by means of *not even close/not even getting close/even quite (fairly) remote/even the opposite,* or *if at all*)?

To answer these questions, we exhaustively extracted from HeTenTen 153 negative constructions, those studied here, followed by *but not Y either* (where the slot marked by *Y* can accommodate any adjectival phrase, noun phrase, or verb phrase) as in examples (10–11) below. (For a description of the extraction procedure, see Appendix C).

(10) … after all **we don't really love**, *but* we *don't really hate either.*
(11) **They were not particularly nice** *but not awful either.*

We further extracted all 8 instances of such negative constructions followed by *not even close/not even getting close/even quite (fairly) remote/even the opposite* as in example (12):

(12) **Not especially a genius idea**, *not even getting close to good.*

We then extracted all 12 instances of such constructions followed by *if at all*, as in example (13):

(13) … without it **it's not really possible to obtain satisfactory results**, *if at all.*

From each list we further removed sequences in which the *X* slot hosts a *non*-positive concept, since such constructions are not tested here, as they are not likely to be interpreted as sarcastic; instead, as shown for such affirmatives, they either convey a compositional (e.g., literal) interpretation or an "ironic praise", the latter conveying a much weaker message in terms of sarcasm (see, e.g., Dews and Winner, 1995; Goldenberg, 2011; Schwoebel et al., 2000; see also Section 4 above). Of the 153 negative constructions studied here, followed by *but not Y either*, 116 instances involving positive concepts (evaluated as such by the first author and an assistant) survived this filtering stage. All 8 instances of such constructions, followed by *not even close/not even getting close/even quite (fairly) remote/even the opposite*, and additional 10 (out of 12) instances followed by *if at all*, also hosting a positive concept (evaluated as such by the first author and an assistant), survived this filtering stage.

Before running the analyses, comparing the number of all these instances of negative constructions (involving either *but not Y either, not even close,* or *if at all*), we had to rate this set of stimuli in terms of degree of nonliteralness (sarcastic vs. literal) to confirm:

a) that the speaker, using *but not Y either* following these negative constructions, wishes to signal that she uses them in order to convey a nondefault compositional (often literal) message, instead of a default, unintended sarcastic alternative; and

b) that if absolute negators, such as *not even close/not even getting close/even quite(fairly) remote/even the opposite* or *if at all*, do follow the negative constructions studied here, then it is because the speaker intends to communicate a nondefault compositional/literal message rather than a default, sarcastic one.

To this end, we presented 3 judges, versed in the field of sarcasm, with the list of all 134 instances of negative constructions, embedded in their natural context, in which, however, all types of cues (116 instances of *but not Y either*, 8 instances of *not even close/not even getting close/even quite (fairly) remote/even the opposite*, and 10 instances *if at all*) were *deliberately* omitted beforehand, so that they would not affect judgments (for a similar methodology, see Kreuz and Caucci, 2007). In the absence of these cues, the judges were asked to decide whether the negative constructions studied here are sarcastically or literally intended.

Results show that significantly more instances of the negative constructions studied here, which, prior to the omission of the cues, were followed by *but not Y either*, were rated as sarcastic rather than literal. Specifically, 67.2% of these 116 cases (78/116 = 67.2%) were judged as sarcastic by at least 2 judges, whereas only 32.8% (38/116 = 32.8%) were judged as literal by at least 2 judges, binomial test, $p < 0.001$. (Only 12 items were judged as literal by all 3 judges). As for the 8 items, which, prior to the omission of the cues, were followed by *not even close/not even getting close/even quite (fairly) remote/even the opposite*, 7 (7/8 = 87.5%) were judged as sarcastic by at least 2 judges, whereas the remaining 1 (1/8 = 12.5%) was judged as literal by all 3 judges. Clearly, of the 8 cases, originally followed by absolute negators, there are more cases judged as sarcastic than literal. However, the difference is statistically marginal, binomial test, $p = 0.07$. As for the 10 items originally followed by *if at all*, 1 item could not be judged by any of the judges, due to lack of sufficient context; 6 of the remaining 9 items (6/9 = 66.6%) were judged as sarcastic by at least 2 judges, whereas the rest 3 (3/9 = 33.3%) were judged as literal by at least 2 judges. Although there are more cases judged as sarcastic than literal, the difference is not statistically significant, binomial test, $p = 0.51$.

These results suggest that *but not Y either* applies when a speaker wishes to convey a literal message (using the negative constructions studied here), but suspects that her addressee would interpret her message sarcastically *by default*, as did our 3 judges. In the case of the very few items, in which the negative constructions are followed by absolute negators, the results are not conclusive. Note that, while we cannot provide any statistically significant support for the assumption that these negative constructions (followed by absolute negators) are necessarily literally intended, we can neither provide any support for the alternative that these constructions are sarcastically intended. And, again, although it was always the case that these items were judged as more sarcastic than literal, their small numbers might have rendered these differences marginally significant or even insignificant.

Still, compared to the small number of cases followed by absolute negators, the number of cases of the negative constructions followed by a cue (*but not Y either*) designated to reject their default sarcastic interpretation, is larger and amounts to 116 cases. Taking into account the prevalence, in HeTenTen, of the sequence *but not Y either*, compared to *not even close/not even getting close/even quite (fairly) remote/even the opposite* or *if at all* (for a detailed description of the procedure by which we determined the entire population of each cue in the corpus, see Appendix C), results show that there are significantly more cases of *but not Y either*, following the negative constructions studied here (116/2435 = 4.76%), than *if at all*, following such negative constructions (9/17382 = 0.05%), χ^2(for homogeneity) = 749.10, $p < 0.001$. There are also significantly more cases of the sequence *but not Y either*, following the negative constructions studied here (116/2435 = 4.76%) than cases of *not even close/not even getting close/even quite (fairly) remote/even the opposite*, following such negative constructions (8/682 = 1.17%), χ^2(for homogeneity) = 18.00, $p < 0.001$. Apart from attesting to the superfluousness of intensification of such constructions (by absolute negators), such results substantiate the defaultness of the sarcastic interpretation of the negative constructions studied here by indicating the need to signal nondefault counterparts, when intended.

All in all, these results support Giora et al.'s (2015b) Defaultness Hypothesis, showing that the negative constructions studied here are interpreted as sarcastic by default. This default sarcastic interpretation could possibly be the reason why a speaker who wishes to use these constructions literally, would indicate, by adding *but not Y either*, that their default unintended interpretation needs to be rejected in favor of an intended nondefault alternative. Note, further, that speakers scarcely use intensifiers, such as *not even close/not even getting close/even quite (fairly) remote/even the opposite* or *if at all*, following such constructions; being sarcastic and therefore blunt by default, they need no further intensification. Indeed, in such cases, the resulting sequence is informationally redundant. In contrast, these intensifiers are in order when they follow such negative constructions when intended literally.

6.2. 'Not seriously' (Hebrew *stæːm*): rejecting default interpretations of affirmatives and negatives while inviting nondefault counterparts

In Modern colloquial Hebrew, there is a procedural discourse cue, pronounced [*stæːm*], which means 'not seriously'/'just kidding'. As shown by Ziv (2013), this procedural *stæːm* rejects what-is-said interpretations (termed here "default" interpretations) in favor of truth-compatible counterparts (termed here "nondefault"). Relevant to our discussion, however, is Ziv's additional suggestion that *stæːm* further telegraphs to the addressee that (prior) information, which *stæːm* refers to, is said in jest and is intended sarcastically, rejecting the default interpretation of what is said while projecting some humorous criticism.

Our corpus search reveals that, at times, this *stæːm* is further strengthened by additional explicit cues such as *just kidding* (see example 14), *LOL* ('laugh out loud'; see example 15), or *and now seriously* (see example 16), all rejecting the default interpretation while prompting an intended, nondefault alternative. Rejecting the default interpretation, while inviting a nondefault (here sarcastic) counterpart, in fact discloses speakers' awareness of degree of defaultness (alongside their intent, so it seems, to mislead comprehenders down the garden path in order to eventually amuse them):

(14) Maybe they got some money from Ruby Bennett so that they write about him in the newspaper …
      stæːm just kidding.
(15) … the plan is … to expose the banking industry in Israel to competition … *stæːm! LOL!* Of course you
      don't do such stuff…
(16) Oh Meir Meir … you captivated me ever since "Russian Novel", and since then I am a dedicated fan of yours,
      I hung a poster of yours above my bed and I wear your number 1 shirt (*stæːm*) *and now seriously*…

To test the prediction of the Defaultness Hypothesis, expecting the procedural cue − *stæːm* − to reject unintended default interpretations, mostly of affirmatives, (whether literal or sarcastic), while inviting a nondefault counterpart (whether sarcastic or literal), we extracted from HeTenTen *all* instances of the procedural *stæːm*. (For a description of the procedure by which we compiled the relevant list of *stæːm* from the corpus, see Appendix D). In all, 239 instances of an interpretable *stæːm*, of which only 1 item was in the form of the negative constructions studied here (see example 17 below, in bold):

(17) I must say that this picture **doesn't really do you any justice**… *stæːm* the truth is that you look there really great!

Inspection of the 239 interpretable instances followed by *stæːm* (run by the first author and an assistant, both versed in the field of sarcasm), reveals that the default sarcastic interpretation (related to looking ugly) of example (17) — *this picture doesn't really do you any justice* — is rejected in favor of the nondefault literal alternative — *you look really great* — which is immediately made explicit. The interpretation of the remaining 238 utterances (mostly in the affirmative, and when in the negative − not the kind of negative constructions studied here) is literal (see examples 14−16 above). This literal interpretation is rejected by *stæːm*, while inviting a nondefault sarcastic alternative, sometimes made explicit by an opposite which follows (see example 15 above).

Taken together, then, these corpus-based results support the Defaultness Hypothesis. They show that, as predicted, cueing by 'not seriously' − the procedural Hebrew cue *stæːm* − invites nondefault interpretations via rejecting unintended default alternatives. This, in fact, is true of *all* the cases studied here: All our items reject unintended default interpretations while inducing nondefault counterparts − the 238 default literal affirmatives and negatives, encouraged to be interpreted sarcastically, and the *single* default negative sarcasm (exemplifying the constructions studied here), explicitly encouraged to be interpreted literally.

In sum, we have introduced two cues which follow both negative and affirmative utterances, aimed at rejecting their default interpretation, while prompting their nondefault alternative:

a) *But not Y either* rejects the default sarcastic interpretation of the negative constructions studied here by explicitly
    stating it, thus inviting a nondefault (literal) counterpart (see Section 6.1).
b) *Stæːm* rejects the default literal interpretation of (mostly) affirmative utterances, while prompting their nondefault
    sarcastic counterparts. When this cue (very rarely) applies to negative constructions, such as those studied here,
    it invites a nondefault (literal) alternative (as shown by example 17).

7. Conclusions

In this study, we put the Defaultness Hypothesis (Giora et al., 2015b) to corpus-based test. The attempt is to provide converging usage-based evidence to online experimental findings, demonstrating the processing speed superiority of default over nondefault interpretations, as predicted by the Defaultness Hypothesis. Specifically, Giora et al. (2015b) show that default affirmative literalness (*s/he is the most alert person around*, meaning 's/he is very alert') is processed faster than both, nondefault negative literalness (*s/he is not the most alert person around*, meaning 's/he is alert but others are more alert than her') and nondefault affirmative sarcasm (*s/he is the most alert person around*, meaning 's/he is far from being alert'). More importantly, default negative sarcasm (*s/he is not the most alert person around*, meaning 's/he is far from being alert') is processed faster than both, nondefault negative literalness and nondefault affirmative sarcasm (Giora et al., 2015b: Exp. 2). Given the defaultness of affirmative literalness and negative sarcasm, the Defaultness Hypothesis further predicts that nondefault counterparts (i.e., nondefault negative literalness and nondefault affirmative sarcasm) will rely heavily on cueing for their derivation. Hence, when intended, speakers will cue nondefaultness while rejecting defaultness (as unintended).

We start, though, by re-establishing, usage-wise, the defaultness of our negative targets, extracted from HeTenTen, by highlighting speakers' sensitivity to their default sarcastic interpretation (see a−b below). We show that

a)   speakers refrain from hosting such negative constructions in linguistic slots intended for weak messages, thus implicitly confirming the bluntness of their default sarcastic interpretation; we further show that

b)   speakers mitigate the blunt message of these sarcastic interpretations by using a metalinguistic mitigator (e.g., *to put it mildly*) rather than an intensifier (*not to be blunt, but*), thereby testifying to their sensitivity to the default harsh message these constructions convey.
We then show that, as predicted (see c—d below), when intended,

c)   speakers prompt the nondefault literal interpretation of these negative constructions (*s/he is not the kindest person around*) by rejecting their default unintended sarcastic interpretation, while using explicit cues, such as *but not Y either* (as in *but not inhumane either*); and finally, we show that

d)   speakers prompt the nondefault sarcastic interpretation of affirmative utterances by explicitly rejecting their default compositional (often literal) interpretation, using cues such as *stæːm* (Hebrew for 'not seriously') or *just kidding*.

In sum, we come up with converging, corpus-based evidence, attesting to speakers' sensitivity to default interpretations, which, in turn, allows them to reject these interpretations while prompting nondefault counterparts (when intended). Such results provide further support for the Defaultness Hypothesis, predicting the need to cue nondefault (rather than default) interpretations (when intended).

Funding

Acknowledgements

Appendix A

HeTenTen is a web-corpus compiled using a web-crawler. It contains approximately 1.2×10^6 web documents which were mined, filtered, and processed using a generic algorithm suggested by Baroni et al. (2009). Crucially, the corpus was crawled in a way that would not allow it to be biased by topic while yet would cover a wide range of language varieties as represented over the web, including semi-spoken languages.

Each word of HeTenTen (i.e., 'surface form') was Part-of-Speech-tagged and morphologically annotated for additional morphological features (see Adler, 2007), such as gender, number, affixes, etc.[12] The corpus can be queried by using an extended version of a Corpus Querying Language (CQL)[13] (Kilgarriff et al., 2014),[14] which enables users to retrieve lines whose patterns are defined by specifying sequences of token and sub-token-level features. Due to the morphological annotation (Adler, 2007), the queries can be defined based on many features which address the rich morphology of Hebrew. Additionally, upon request, a wider context can be suggested beyond the line/sentence level, and there is always a pointer to the URL from which the web page was crawled.

Appendix B

In the extraction process of *to put it mildly* and *(not) to be (too) blunt/rude* (followed or preceded by the negative constructions studied here), we focused on the following sequences which seemed to cover all the possible variants of *to put it mildly* and *(not) to be (too) blunt/rude*:

To put in English translates into *to say* in Hebrew. Two variants of *to say*, both in the infinitive, were used (*le-hagid* and *lomar*); the infinitive forms of these two variants of *to say* are commonly replaced by the verbal forms of the singular or the plural in the future tense. Each of these was either immediately followed by *mildly* or separated from *mildly* by up to 2 words.

---

[12] https://www.sketchengine.co.uk/hetenten-corpus/#Tokenattributes.
[13] http://cwb.sourceforge.net/temp/CQPTutorial.pdf.
[14] http://www.sketchengine.co.uk.

The resulting *to put it mildly* sequences, which were *not* comments on a prior or a proceeding utterance, were further filtered out from the resulting list. We ended up with a list of 872 instances involving the *to put it mildly* comment.

Unlike the *to put it mildly* query, in which the sequences queried were quite constrained, the compilation of the *(not) to be (too) blunt/rude* subset involved extraction of utterances containing the adverbs *bluntly*, *in a blunt manner*, and the adjective *blunt* (put together, 2712 items), which are clearly considerably less constrained queries. This is because *bluntly*, *in a blunt manner*, and the adjective *blunt* are sporadically used (for obvious reasons made clear in Section 5), and therefore have not evolved into any distinct entrenched sequence like the *to put it mildly* variants, which are discursively-motivated, more frequent, and therefore linguistically-entrenched. From the exhaustive—and rather extended—list of 2712 items, containing the adverbs *bluntly*, *in a blunt manner*, and the adjective *blunt*, we further filtered out any item which was *not* a comment on either a prior or a proceeding utterance, ending up with 78 items roughly in the form of *(not) to be (too) blunt/ rude*

From the two exhaustive lists (of *to put it mildly* and *(not) to be (too) blunt/rude*), described in the previous paragraphs, we further extracted all instances of *to put it mildly* and *(not) to be (too) blunt/rude* followed or preceded by the negative constructions studied here.

Appendix C

We extracted from HeTenTen *all* 4831 sequences of *but not Y either*, from which we further extracted this cue, following the negative constructions studied here. The same procedure was applied in the case of the 701 instances of *not even close/not even getting close/even quite (fairly) remote/even the opposite*, and the 19081 instances of *if at all*.

As HeTenTen is not syntactically parsed, the 3 abovementioned lists contain sequences that are *only* superficially identical to the 3 kinds of cues which are of interest to us. Therefore, to determine the general population in the corpus of each of the 3 cues—*but not Y either*, *not even close/not even getting close/even quite (fairly) remote/even the opposite*, and *if at all*—while removing unrelated superficially identical sequences, we followed the ensuing procedures:

But not Y either: Not all 4831 instances of *but not Y either* display a contrast between the utterance preceding *but* and the utterance *Y,* following it. There are quite some instances in which the proposition following *but* is no more than a comment on a prior utterance, as in example (A1):

(A1)   I didn't see *but (I) didn't* look for it *either.*

Such items, although using the same *but not Y either* surface form, are pragmatically different from those which use this form in order to express a contrast. In order to focus on the population of contrastive *but not Y either* sequences, we pseudo-randomly sampled 1000 instances out of the exhaustive list of 4831 instances of *but not Y either*, and manually sorted them into two categories — the contrastive category and the topic-comment-relation category. Results show that 504 instances out of the 1000 sampled instances were of the contrastive type. Extrapolating these results suggests that 2435 cases out of the 4831 instances of *but not Y either* are expected to be of the contrastive type.

Not even close: Not all 701 cases of *not even close/not even getting close/even quite (fairly) remote/even the opposite* serve as absolute negators, but are rather used literally to signal spatial proximity, as in example (A2):

(A2)   Why move to Sarah village? It's *not even close* to dad's and mom's workplace.

Instances such as example (A2) were discarded from this specific list, thus narrowing it to a list of 682 instances of *not even close/not even getting close/even quite (fairly) remote/even the opposite*, each of which serving as an absolute negator, intended to drive the interpretation of a prior utterance towards the opposite extreme end of a scale.

If at all: Not all 19081 cases of *if at all* serve as an absolute negator appositive tag, intended to negate altogether the utterance preceding them. Some cases introduce the antecedent of a conditional (see example A3), some introduce a complement (see example A4), and some are category wideners (see example A5), as in the following examples, respectively:

(A3)   If you are …of the kind of travelers who arrived here to try their luck with surfing for the first time, and *if* you didn't even plan it *at all*, but you are already here …[15]
(A4)   Because we didn't know *if at all* we would go away from here…
(A5)   Please welcome the parade of picture frames which will furnish your home with the family-like look, whether on family day or *if at all.*

---

[15] As a consequence of different word orders in Hebrew and in English, *if* and *at all* in this specific example, translated from Hebrew, do not constitute a single sequence, whereas in Hebrew they do.

As in the case of *but not Y either,* here too, we applied that same filtering procedure to *if at all*: We pseudo-randomly sampled 1000 instances out of the exhaustive list of 19081 instances of *if at all*, from which we carefully filtered out instances of *if at all* serving any purpose other than an absolute negator.

Results show that 911 out of the sampled 1000 instances were of the absolute negator appositive tag type. Extrapolating these results suggests that 17382 out of the 19081 instances of *if at all* are expected to be of the absolute negator type.

Appendix D

As the procedural *stæ:m* is an instance of informal colloquial spoken Modern Hebrew, it is never used in edited texts, and therefore has no formal orthography agreed upon. In informal corpora it would be transcribed as $st(a)_nm$, n running from 1 upwards.

In order to have as many instances of *stæ:m* as possible, we queried $st(a)_nm$s, in which n runs from 1 to 20. However, starting from $n = 13$, the queries returned no results. It is not impossible, however, that instances of $st(a)_nm$ in which $12 < n$ do exist outside of HeTenTen.

The raw data comprised of 400 instances of $st(a)_nm$ $n = 1–12$ from which we further discarded the following instances of *stæ:m*, ending up with 239 interpretable items:

a) 100 cases of $st(a)_nm$, in which the speaker mistakenly used the procedural *stæ:m* in place of the conceptual *stām* which has a fixed orthography and is written with no *a*'s at all. (Most of these misprints are $st(a)_nm$, $n = 1$).
b) 28 cases of $st(a)_nm$ serving as a name of a person or a TV program.
c) 16 cases of $st(a)_nm$ in which *stæ:m* was not produced by the speaker herself, but was rather a quote of another speaker.
d) 15 cases of $st(a)_nm$ in which *stæ:m* seemed like a detached fragment.
e) 2 cases of $st(a)_nm$ attempting to define *stæ:m* and its terms of use.

References

Adler, Meni, 2007. Hebrew Morphological Disambiguation: an Unsupervised Stochastic Word-based Approach. unpublished PhD dissertation. Ben-Gurion University of the Negev, Israel.

Aijmer, Karin, 2013. Understanding Pragmatics Markers. Edinburgh University Press, Edinburgh.

Ariel, Mira, 2008. Pragmatics and Grammar. Cambridge University Press, Cambridge.

Attardo, Salvatore, 2000. Irony markers and functions: towards a goal-oriented theory of irony and its processing. Rask Int. J. Lang. Commun. 12 (1), 3–20.

Bardenstein, Ruti, 2019. Grammaticalization Paths of Rectification Constructions. unpublished PhD dissertation. Tel-Aviv University, Israel.

Baroni, Marco, Bernardini, Silvia, Ferraresi, Adriano, Zanchetta, Eros, 2009. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. Lang. Resour. Eval. 43 (3), 209–226.

Booth, Wayne C., 1974. A Rhetoric of Irony. University of Chicago Press, Chicago, IL.

Brown, Penelope, Levinson, Stephan C., 1987. Politeness: Some Universals in Language Usage. Cambridge University Press, New-York, NY.

Dews, Shelly, Winner, Ellen, 1995. Muting the meaning: a social function of irony. Metaphor Symbol 10 (1), 3–19.

Du Bois, John W., 2003. Argument structure: grammar in use. In: Du Bois, J.W., Kumpf, L.E., Ashby, W.J. (Eds.), Preferred Argument Structure: Grammar as Architecture for Function. John Benjamins, Amsterdam, pp. 11–60.

Du Bois, John W., 2014. Towards a dialogic syntax. Cognit. Ling. 25 (3), 359–410.

Filik, Ruth, Howman, Hannah, Ralph-Nearman, Christina, Giora, Rachel, 2018. The role of defaultness in sarcasm interpretation: evidence from eye-tracking during reading. Metaphor Symbol 33 (3), 148–162.

Filik, Ruth, Ţurcan, Alexandra, Thompson, Dominic, Harvey, Nicole, Davies, Harriet, Turner, Amelia, 2016. Sarcasm and emoticons: comprehension and emotional impact. Q. J. Exp. Psychol. 69 (11), 2130–2146.

Gibbs, Raymond W., 1994. The Poetics of Mind: Figurative Thought, Language, and Understanding. Cambridge University Press, New York, NY.

Gibbs, Raymond W., 2000. Irony in talk among friends. Metaphor Symbol 15 (1–2), 5–27.

Giora, Rachel, 1997. Understanding figurative and literal language: the graded salience hypothesis. Cognit. Ling. 8 (3), 183–206.

Giora, Rachel, 2003. On Our Mind: Salience, Context, and Figurative Language. Oxford University Press, New York, NY.

Giora, Rachel, 2018. How defaultness shapes our language production: a usage-based study of discoursal resonance with default interpretations of metaphor and sarcasm. In: Barnden, J., Gargett, A. (Eds.), Production of Figurative Language. John Benjamins. Submitted.

Giora, Rachel, Cholev, Adi, Fein, Ofer, Peleg, Orna, 2018a. On the superiority of defaultness: hemispheric perspectives of processing negative and affirmative sarcasm. Metaphor Symbol 33 (3), 163–174.

Giora, Rachel, Drucker, Ari, Fein, Ofer, Mendelson, Itamar, 2015a. Default sarcastic interpretations: on the priority of nonsalient interpretations. Discourse Process. 52 (3), 173–200.

Giora, Rachel, Givoni, Shir, Becker, Israela, 2018. How defaultness affects text production: a corpus-based study of default sarcasm. In: Athanasiadou, A., Colston, H. (Eds.), The Diversity of Irony, CLR Book Series. De Gruyter Mouton. Submitted.

Giora, Rachel, Givoni, Shir, Fein, Ofer, 2015b. Defaultness reigns: the case of sarcasm. Metaphor Symbol 30 (4), 290–313.

Giora, Rachel, Givoni, Shir, Heruti, Vered, Fein, Ofer, 2017. The role of defaultness in affecting pleasure: the optimal innovation hypothesis revisited. Metaphor Symbol 32 (1), 1–18.

Giora, Rachel, Jaffe, Inbal, Becker, Israela, Fein, Ofer, 2018b. Strongly mitigating a highly positive concept: the case of default sarcastic interpretations. Rev. Cognit. Ling. 6 (1), 19–47.

Giora, Rachel, Livnat, Elad, Fein, Ofer, Barnea, Anat, Zeiman, Rakefet, Berger, Iddo, 2013. Negation generates nonliteral interpretations by default. Metaphor Symbol 28 (2), 89–115.

Givoni, Shir, Giora, Rachel, Bergerbest, Dafna, 2013. How speakers alert addressees to multiple meanings. J. Pragmat. 48 (1), 29–40.

Goldenberg, Dolly, 2011. Default Ironic Interpretation. Unpublished ms.. Tel Aviv University, Israel.

Grice, Herbert Paul, 1975. Logic and conversation. In: Cole, P., Morgan, J.L. (Eds.), Syntax and Semantics: Speech Acts, vol. 3. Academic Press, New York, NY, pp. 41–58.

Haiman, John, 1998. Talk Is Cheap: Sarcasm, Alienation, and the Evolution of Language. Oxford University Press, New York, NY.

Hary, Benjamin, Izre'el, Shlomo, 2003. The preparatory model of the Corpus of Spoken Israeli Hebrew (CoSIH). In: Hary, B. (Ed.), Corpus Linguistics and Modern Hebrew: Towards the Compilation of the Corpus of Spoken Israeli Hebrew (CoSIH). Tel Aviv University, The Chaim Rosenberg School of Jewish Studies, Tel Aviv, pp. 189–219.

Horn, Laurence R., 1989. A Natural History of Negation. University of Chicago Press, Chicago, IL.

Katz, Albert N., Ferretti, Todd R., 2003. Reading proverbs in context: the role of explicit markers. Discourse Process. 36 (1), 19–46.

Kilgarriff, Adam, Baisa, Vít, Bušta, Jan, Jakubíček, Miloš, Kovář, Vojtěch, Michelfeit, Jan, et al., 2014. The sketch engine: ten years on. Lexicography 1 (1), 7–36.

Kreuz, Roger J., 1996. The use of verbal irony: cues and constraints. In: Mio, J.S., Katz, A.N. (Eds.), Metaphor: Implications and Applications. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 23–38.

Kreuz, Roger J., Caucci, Gina M., 2007. Lexical influences on the perception of sarcasm. In: Proceedings of the Workshop on Computational Approaches to Figurative Language. Rochester, NY, 26 April.

Muecke, Douglas Colin, 1978. Irony markers. Poetics 7 (4), 363–375.

Oh, Sun-Young, 2000. *Actually* and *in fact* in American English: a data-based analysis. Engl. Lang. Ling. 4 (2), 243–268.

Ran, Yongping, 2015. Metapragmatic negation as a rapport-oriented mitigating device. J. Pragmat. 84, 190–203.

Schwenter, Scott A., Traugott, Elizabeth Closs, 2000. Invoking scalarity: the development of *in fact*. J. Hist. Pragmat. 1 (1), 7–25.

Schwoebel, John, Dews, Shelly, Winner, Ellen, Srinivas, Kavitha, 2000. Obligatory processing of the literal meaning of ironic utterances: further evidence. Metaphor Symbol 15 (1–2), 47–61.

Sperber, Dan, 1984. Verbal irony: pretense or echoic mention? J. Exp. Psychol. Gen. 113 (1), 130–136.

Sulis, Emilio, Farías, Hernández, Irazú, Delia, Rosso, Paolo, Patti, Viviana, Ruffo, Giancarlo, 2016. Figurative messages and affect in Twitter: differences between #irony, #sarcasm and #not. Knowl. Based Syst. 108, 132–143.

Tannen, Deborah, 1984. Conversational Style. Analyzing Talk Among Friends. Ablex Publishing Corporation, Norwood, NJ.

Thompson, Dominic, Filik, Ruth, 2016. Sarcasm in written communication: emoticons are efficient markers of intention. J. Comp. Mediat. Commun. 21 (2), 105–120.

Traugott, Elizabeth Closs, Dasher, Richard B., 2002. Regularity in Semantic Change. Cambridge University Press, Cambridge.

Veale, Tony, 2018. The "default" in our stars: signposting non-sefaultness in ironic discourse. Metaphor Symbol 33 (3), 175–184.

Wasow, Thomas, 1997. End-weight from the speaker's perspective. J. Psycholinguist. Res. 26 (3), 347–361.

Ziv, Yael, 2013. 'Staam': Šmirat ʕikviyut ba-siax (Hebrew 'staam': maintaining consistency in discourse). In: Florentin, M. (Ed.), Mexkarim ba-ʕivrit ha-xadasah u-ve-mekoroteha le-zexer Shaul Aloni (Studies in Modern Hebrew and its Origins in memory of Shaul Aloni). The Academy of the Hebrew Language, Jerusalem, pp. 151–159 (in Hebrew).

Israela Becker is a PhD candidate in the School of Philosophy, Linguistics and Science Studies at Tel-Aviv University. Her main research interests lie in the psycholinguistics and pragmatics of negation. Apart from a Major and a Master's degree in linguistics, she holds a PhD in chemistry.

Rachel Giora is Professor of Linguistics and Psycholinguistics at Tel-Aviv University. Her research areas include cognitive aspects of coherence, and women and language. As of 1997, her work has focused on experimentally testing the Graded Salience Hypothesis and the Defaultness Hypothesis (2015, 2017), featuring the psycholinguistics and neurolinguistics of non/default, non/figurative language, context effects, Optimal Innovations, aesthetic pleasure, and discourse negation. She has published over 130 articles, a book titled *On Our Mind* (OUP, 2003), coedited a 6 volume series with Patrick Hanks on figurative language (2011), and a volume coedited with Michael Haugh, titled *Doing Intercultural Pragmatics* (2017).