Grice, Machine Head and the problem of overexpressed premises

MICHAEL J. HOPPMANN
Northeastern University
m.hoppmann@neu.edu

The natural language phenomenon of "unexpressed premises" (UP) – statements that protagonists do not explicitly utter, but to which they are nevertheless committed – is well documented. This paper argues that its contrary opposite "overexpressed premises" (OP) – statements that protagonists explicitly utters, but to which they are nevertheless not committed – frequently occurs in the usage of dialectical irony (as illustrated in Machine Head lyrics), and that reasonable OP interpretations require additional reconstructive tools beyond Gricean Maxims.

KEYWORDS: Grice, irony, maxims, overexpressed premises, reasonableness, reconstruction, unexpressed premises

1. INTRODUCTION

This paper uses a Gricean perspective to take a closer look at the reasonable interpretation of Socratic irony. Utilizing a Gricean, Post-Gricean or Neo-Gricean approach to understanding irony is an extremely well-trodden path that starts with Grice himself. He presents irony as the first example for the flouting of the first maxim of quality in his seminal paper (Grice 1989, 34). Trying to even list, let alone summarize or critique, the books and papers that have followed, extended or opposed him on that path would be a herculean task – one that I have no intention of undertaking in this essay. Instead I would like to concentrate on a particular subfield and shift the focus in two distinct regards to visit a part of this path that I believe still deserves additional attention.

The first shift regards the type of irony under scrutiny. Grice himself uses an example of contrary irony ("X is a fine friend" said about a known traitor, Grice 1989, 34) in his paper. Others have since added discussions of a number of related types of irony – most of which, are based on a contrary opposition between what is literally expressed and what is meant by the speaker. This type of irony is a fascinating

communicative phenomenon, but it is not the only type that deserves our attention. Leaning closer on the definitions of irony provided by Burke and Lausberg as well as the examples given by Plato, I will focus on a different kind of verbal irony that does not necessarily contain a contrary opposite at its core; namely dialectical or Socratic irony.

The second shift concerns the aim of studying the relationship between Gricean maxims and irony. Most recent works on Grice and irony seem to give a strong emphasis to the fundamental theoretical understanding of irony in communication. My aim here is more modest and applied. I believe one of the most beautiful aspects of the Gricean maxims for the argumentation theorist is their usefulness as reconstructive tools, in cases in which the interpretation of a statement is disputed and a protagonist[1] denies responsibility for statements that an antagonist[2] attributes to him. My question in this essay is thus not one of philosophical sense-making but of reasonable expectations about responsibility and deniability.

Given this aim, the current essay naturally falls into three main parts: a. the nature of Socratic irony, its markers and relationship to other figures, illustrated by examples of Socratic irony, b. the Gricean maxims as tools for reasonable reconstruction of contested meaning, and c. the application of the Gricean maxims to the special case of Socratic irony. Unfortunately, the third of these parts will be mainly aporetic, showing the limitations of an otherwise helpful tool without a clear solution to the problems posed.

Before approaching these three main sections, one important terminological problem deserves to be addressed. Colloquial language is ill-equipped for the discussion of some linguistic challenges, the main case in point here being the meaning of "say". This has led Grice and many scholars since him to introduce terminological distinctions between "say" in "his favored sense" (Grice 1989, 25, 33; Dynel 2018, 34ff.) and "say" in a more general sense, as well as between "saying", "making as if to say" and related terms such as "asserting", "expressing", "stating" etc. Very few of these distinctions correlate with ordinary language use, and significant parts of recent scholarship have been dedicated to clarifying or criticizing previous distinctions. Part of this discourse is helpful for the understanding of Grice and Gricean, post-Gricean, and neo-Gricean scholarship, but this essay is no attempt to make a contribution to either of these fields. So rather than entering this complex theoretical discourse, I will use simple indexing to distinguish between "saying$_L$" for literal statements independent of the modality of their expression or the kind of speech act employed, and "saying$_R$" for

[1] In this paper referred to as P and with male pronouns.
[2] In this paper referred to as A and with female pronouns.

the (implied or explicit) meaning of a statement for which the speaker or writer is reasonably expected to take responsibility. Where necessary this indexing will also be extended to "statement_L/statement_R" and similar terms.

2. SOCRATIC IRONY

"Irony" is one of the most notoriously ambiguous concepts in rhetorical theory, rivalled only by a handful of other terms ("*topos*", "*ethos*", "*parrhesia*" or "*enthymeme*" come to mind) in the race for largest number of contradictory definitions provided in the literature. Distinguishing between these definitions and analyzing their relationship to each other is a noble undertaking – but not one than can realistically be attempted by a single paper – let alone one with a different main aim.[3] Providing a rough approximation of the location of Socratic irony within the plentitude of ironies, might nevertheless be useful for an understanding of the concept. To do this I will briefly introduce a set of four imperfectly concentric circles of irony. 'Imperfectly concentric' because while I think there is a good argument to be made about the order of vagueness and conceptual sizes of the respective circles, some of the following meanings include each other, while others can (also) be read as exclusive to one another.

(1) Irony in the widest sense – popular irony – refers to likely the most frequent understanding of irony. Irony in this sense, the sense of Alanis Morissette and countless internet memes, refers to a variety of conceptual fields, including "surprising", "coincidental" or "unfortunate", or even "just" (in the sense of payback of karma). The popular usage of the term in this sense is also one of the causes of some terminological and empirical challenges involved in the study of irony (Dynel 2018, 20ff., 137ff.; Simpson 2011, 36ff.; Burgers et al. 2011, 187ff.).[4]

(2) Irony in the wider sense – existential irony – is the irony of Hegel, Schlegel, Kierkegaard. It is better defined than popular irony, but like the latter, refers to concepts that lay mainly outside of the realm of rhetoric and communicative phenomena (Behler 1998, 607ff.; Braungart 2010, 323ff).

(3) Irony in the wide sense – situational irony – differs from the popular and existential irony in that it clearly refers to a communicative

---

[3] Some excellent introductions into the various meanings of 'irony' are provided by Muecke 1969, 64ff.; Ibid. 1970, 7ff.; Behler 1998, 599ff.; Dynel 2018, 157ff.

[4] Fogelin presents a strong case in favour of some cases of popular irony, namely poetic justice, structurally resembling irony in the narrow sense, thus providing a connection between the popular usage and the more theoretical terminology (2011, 22f.).

situation. Unlike verbal irony, the main communication concerned is not the one between protagonist and antagonist, but instead happens between a (real or imagined) author of the situation and his or her audience. Situational irony is thus marked by a tension between partially ignorant agents and a better-informed audience that observes the irony in the statements and actions of the agents (see also Fogelin 2011, 21ff.).

(4) Irony in the narrow sense – verbal irony – is the irony with which this essay is concerned. This type of irony describes a figure of speech or form of communication that is (like other tropes) marked by a contrast between what is literally expressed ($said_L$) and what is meant ($said_R$). The most common form of verbal irony is the contrary irony of the classical textbooks. Anaximenes describes it as "*calling things by the opposite of their real names*" (Rhet. ad Alex. 1434a, trans. H. Rackham) and Quintilian explains this irony, or *illusio*, as "*the type in which meaning and the words are contrary*" (Quint. Inst. Orat. VIII, 6, 54, trans. Russell). This is also the type of irony Grice is thinking of in his traitor example. Most modern definitions of verbal irony have a similar type in mind and centre around two necessary (but not always jointly sufficient) markers of irony: 1. a contrast between $meaning_L$ and $meaning_R$ – the marker of figurative language in general and 2. a relationship of contrary opposition between $meaning_L$ and $meaning_R$.[5] These core qualities are sometimes supplemented by additional markers, such as delivery clues in Quintilian (Inst. Orat. VIII, 6, 54), stylistic clues or internal conflicts in Booth (1974, 49ff.), or evaluative form and negatively evaluative implicature in Burgers, van Mulken & Schellens and Dynel respectively (Burgers et al. 2011, 189; Burgers et al. 2012, 293; Dynel 2018, 106ff.).[6]

While contrary irony is certainly an important variety of verbal irony, not all types of verbal irony are captured by the markers above. Notably, and most importantly for the purpose if this essay, the second core marker does not apply to another type of irony, Socratic or Dialectical irony. This type of irony is slightly harder to define, but clearly referred to in the works of Kenneth Burke and Heinrich

---

[5] Frequently these authors refer to "opposition" rather than "contrary opposition" as the relevant second marker. Since other types of opposition – namely contradictory oppositions – would reduce the second marker to the logical core of the first marker, one can only assume that the contrary opposite is meant. See also Fogelins 2011, 9ff.

[6] Not all types of verbal irony fit neatly into this summary. See for example Dynel's forth type, surrealistic irony, that seems to work without a contrary opposition between $meaning_L$ and $meaning_R$. (Dynel 2018, 171ff.) Compare also Simpson 2011, 40ff.)

Lausberg. Burke famously equates irony with the dialectic and describes it as:

> *Hence, from the standpoint of this total form (this 'perspective of perspectives'), none of the participating 'sub-perspectives' can be treated as either precisely right or precisely wrong. They are all voices, or personalities, or positions, integrally affecting one another. When the dialectic is properly formed, they are the number of characters needed to produce the total development. […] True irony, humble irony, is based upon a sense of fundamental kinship with the enemy, as one needs him, is indebted to him, is not merely outside him as an observer but contains him within, being consubstantial with him.* (Burke 1941, 432ff.)

Lausberg offers a similar explanation. In his *Elemente der Literarischen Rhetorik* he writes "*Die simulatio besteht in der meist affektisch provozierenden, manchmal auch sich emphatisch harmlos gebenden (also die Wirkungs-Absicht dissimulierenden) positiven Vertretung der Meinung des Parteigegners.*" (Lausberg 1949 §429) He further develops this description in the *Handbuch der Literarischen Rhetorik*:

> "Die Ironie ist der Ausdruck einer Sache durch ein deren Gegenteil bezeichnendes Wort. Sie ist eine Waffe der Parteilichkeit: der Redner ist sich der Überzeugungskraft seiner eigenen Partei sowie der Sympathie des Publikums so sicher, daß er […] die lexikalische Wertskala des Gegners verwendet und deren Unwahrheit durch den (...) Kontext evident werden läßt." (Lausberg 1960, §582).

While this later definition also references a relationship of opposition ("deren Gegenteil"), the emphasis in both descriptions lays on the invocation of the opinion of the opponent ("Meinung des Parteigegners") or the lexical values of the opponent ("die lexikalische Wertskala des Gegners"). In other words, Lausberg's and Burke's understanding of irony coalesce in the idea of the invocation of the opinion or the terminology of the antagonist (or a third party), and its incorporation into the protagonist's message. This usage of the other's voice or perspective thus becomes the defining quality of Socratic irony.

Several decades after Burke and Lausberg, and in apparent impudence of the former, Dan Sperber and Deidre Wilson develop a very similar concept of irony (Sperber & Wilson 1981; Ibid. 1995; Wilson & Sperber 2015). Sperber and Wilson understand (Socratic) irony as and echotic mention of the standpoint of a third party (Sperber & Wilson 1981, 306ff.). Their distinction between the 'usage' of a standpoint in literal (and some figurative) language versus the 'mentioning' of a standpoint without explicit attribution to the referenced party is very useful for the understanding of the reduced

burden of proof that the protagonist assumes for an ironic utterance. The main difference between Sperber and Wilson's concept of echotic irony and Burke's and Lausberg's explanations seems to lie in the formers' limitations to 'thoughts' (Wilson & Sperber 2015, 125) or 'contents' (Ibid, 131) as opposed to the option of the invocation of either a thought or a diction or both in the latter. This distinction is of particular importance in examples two and three below.

3. EXAMPLES OF SOCRATIC IRONY

Since Socratic irony is less extensively discussed in the literature and its definition does not neatly align with the traditional 'meaning$_R$ is contrary to meaning$_L$' structure, I will provide three brief examples of Socratic irony, which illustrate its core qualities and at the same time distinguish it from contrary irony. They also exemplify how irony can be used in very serious contexts and is not necessarily accompanied by a humorous or lighthearted context. To cover a wide breath of discourse types, the first example is taken from a Platonic dialogue, the second from a statement about a police shooting victim and the third from the lyrics of a heavy metal band.

(1) Plato's Socrates states in the Gorgias:

> I am certain that whenever you agree with me in any view that my soul takes, this must be the very truth. For I conceive that whoever would sufficiently test a soul as to rectitude of life or the reverse should go to work with three things which are all in your possession – knowledge, goodwill, and frankness. I meet with many people who are unable to test me, because they are not wise as you are; while others, though wise, are unwilling to tell me the truth, because they do not care for me as you do; and our two visitors here, Gorgias and Polus, though wise and friendly to me; are more lacking in frankness and inclined to bashfulness than they should be; nay, it must be so, when they have carried modesty to such a point that each of them can bring himself, out of sheer modesty, to contradict himself in face of a large company, and that on questions of the greatest importance. But you have all of these qualities which the rest of them lack: you have had a sound education, as many here in Athens will agree; and you are well disposed to me. (Gorg. 486e, trans. W.R.M. Lamb)

Interpreting Socrates literally here does not lead to immediate contradiction. The co- and context of this section strongly suggests however, that Socrates does not consider Callicles a perfect model of highest wisdom, greatest goodwill towards Socrates and strongest frankness. Instead he is clearly speaking ironically. Yet, if the textbook definition of contrary irony were to be applied to this text, meaning$_R$ of Socrates words should be claiming of Callicles that he is the perfect

model of highest folly, greatest malevolence and strongest flattery or deception. This reading is certainly possible, but it doesn't match the context very well either. Instead it appears that Socrates is trying to echo Callicles' own opinion about himself by invoking his voice against him. If he thinks so highly of himself then he should be able to serve as the litmus test of truth, and if his claims will be rebutted in dialogue then that rebuttal will stand firmly.

(2) Terence Crutcher was shot by a police officer in 2016. The incident was filmed by a police helicopter and the helicopter police can be overheard calling Terence Crutcher a "big bad dude". Reacting to his passing, Dr. Tiffany Crutcher commented on the death of her unarmed twin brother by the hands of the police with "*That big bad dude was enrolled at Tulsa Community College. He just wanted to make us proud. That big bad dude loved God.*" (The Guardian 2016) Her statement is a very clear example of Socratic irony and an illustration that irony by now means it has to be humorous. Taken literally, calling a recently deceased victim of police violence, especially a close family member, a "big bad dude" is certainly offensive. The phrasing lacks sufficient respect for her late brother and would thus at face value be insulting. Contrary irony interpretation cannot heal this inappropriateness, as Dr. Crutcher certainly did not mean to call her brother a "small nice gal" or tried to insinuate that he was not wanting to make his family proud. Instead she is clearly invoking the word of the police officer to show the contrast between the officer's words and her message about her brother.

(3) The final example is taken from the lyrics of bay area metal band Machine Head. In their 2018 song "bastards" they sing "*So give us all your faggots, all your niggas, and your spics - Give us all your Muslims, your so-called terrorists - We'll welcome them with open arms, and put 'em in our mix - We're better off together now, embrace our difference*." The song as a whole is a strong statement in favor of tolerance and inclusiveness, so using heavily laden expletives for some of the groups whose discrimination the song calls out, creates a face level discord. After all, the use of this widely shunned terminology usually serves as a marker of racism. As a result, the tension between message and terminology strongly suggests ironic intent, but once again contrary irony is not the right tool to reconstruct what the band is trying to say_R. Machine Head do not suggest to give them "all your straight, white and non-Hispanic people" nor do they ask to "take away all your gays, and black and Hispanic people", but instead they invoke the voice of the (unnamed) homophobe and xenophobe to create a verbal contrast with their main message.

The examples illustrate the core functioning of Socratic irony. As a rhetorical figure it stands halfway between contrary irony and

prosopopoeia. It differs from the former by not being translatable with the help of a simple inversion of (one axis of) its $meaning_L$, and from the latter in the lack of explicit attribution or citations. The main markers of Socratic irony are 1. the common marker of tropes (a contrast between $meaning_L$ and $meaning_R$) and 2. an implicit reference to the content or wording of a (second or) third party (making it possible to insert a "as you said" or "as they might call it" to translate $meaning_L$ into $meaning_R$). Socratic irony will also frequently share non-verbal or stylistic bonus indicators with other forms of verbal irony as well. Given these basic characteristics, the main challenge in the interpretation of Socratic irony is a reasonable reconstruction that leads to a defensible $meaning_R$ given an instance of $meaning_L$.

4. GRICEAN MAXIMS IN REASONABLE RECONSTRUCTION

Interpreting a discursive partner or opponent in a reasonable manner is an essential component of any reasonable discussion. It features in a number of places within the pragma-dialectical model of a critical discussion, probably most prominently in the third and fifth commandments (and their rule counterparts). The former, the 'standpoint rule' stipulates that "*Attacks on standpoints may not bear on a standpoint that has not actually been put forward by the other party."* (van Eemeren & Grootendorst 2004, 191; van Eemeren 2018, 59) and the latter, the 'unexpressed-premise rule', requires that *"Discussants may not falsely attribute unexpressed premises to the other party, nor disown responsibility for their own unexpressed premises."* (van Eemeren & Grootendorst 2004, 192; van Eemeren 2018, 60). Of these, the third commandment primarily regulates the antagonist's behavior, banning her from misinterpreting the protagonist's statements, whereas the fifth commandment addresses the protagonist's and antagonist's discursive behavior equally, requiring them to not unduly add or subtract from P's commitment set based on his statements (van Eemeren 2018, 63).

Under ideal and cooperative circumstances, discussion partners need no further explicit or dependable rules to guide their interpretations. Instead they can resort to a model order of 1) P stating X, 2) A interpreting X, 3) if necessary, A checking her interpretation with P, and 4) P truthfully confirming or correcting A's interpretation. The pragmatic, linguistic, and psychological rules involved in this kind of interpretation are fascinating, but not the subject of this essay (comp. e.g. van Eemeren & Grootendorst 2004, 95ff.; van Eemeren 2015, 94ff. Ibid. 2018, 89ff.; van Eemeren, F. H., et al. 1993, 37ff.; van Eemeren & Snoeck Henkemans 2017, 43ff.). I want to instead address a situation in which the circumstances are less ideal, and participants require more resilient rules.

There are a number of different reasons why the model order above might not be applicable. Without an attempt at an exhaustive organization of these reasons at least three groups of interactions immediately present themselves:

Group 1 including all kinds of mediated communication. If A is reading a book, watching a televised speech or is the recipient of similar, primarily monological communication, then asking P for confirmation of her interpretation will usually simply not be feasible. If her agreement or disagreement depends on a particular interpretative alternative, then she might require a way to test her initial intuitive interpretation. The principle of charity requires A to interpret P in the strongest reasonable way, but while this is a sensible and well-justified ideal, it does little to guide the choice between competing benevolent interpretations (comp. van Eemeren, F. H., et al. 1993, 49ff.; van Eemeren & Snoeck Henkemans 2017, 65ff.)

Group 2 includes situations in which P and A might indeed be in a dialogical face to face conversation, but P might not be entirely certain of the 'right' interpretation of his statement. It might well be possible that he states X without a very clear understanding of which interpretation of X should be taken. If this is the case, then he cannot confirm or reject any of A's interpretative proposals or requests.

Group 3 includes the countless dialogues in which both participants might well agree to discuss reasonably, and thus pursue a (pragma-)dialectical goal, but at the same time also maintain rhetorical aims that are not completely subordinate to the dialectical goals. Expressed in a soccer analogy, both players might want to play soccer with each other, be familiar with the rules and willing to play fair, but given that they also want to win, they will not go out of their way to volunteer an offside call against their own side, especially if they are not completely certain that their team has indeed committed the offense. Put in more technical terms, while P and A might be committed to the first order rules of reasonableness, the required second and higher order conditions might be imperfectly fulfilled, and the dialogue partners might be willing to test the limits of reasonable strategic maneuvering.

Under any of the above imperfect conditions the imperatives provided by the pragma-dialectical model on the large scale may require additional guidelines for their implementation. What does it mean for a standpoint to "have actually been put forward" and for an unexpressed premise to be present? Put in other words, how does one justify a deviation from a literal interpretation or the leap from statement_L to statement_R?

One way of understanding the relationship between an individual literal statement and the matching content of the

commitment set of the speaker is to think of them as being located on a continuum of speaker responsibility addition and subtraction from what is literally said. One end of this spectrum contains pure unexpressed premises (UPs), statements$_R$ that enter the P's commitment set, even though they have no single statement$_L$ that justifies them. Their reasonable existence can be reconstructed from the context but is not the result of an immediate interpretation of a single utterance. On the other end of this spectrum is located what I want to call, for want of a better term, overexpressed premises (OPs); statements$_L$ that even though uttered, do not enter P's commitment set and do not produce a substitution. Thinking of reasonable reconstructions and interpretations in this manner would then allow us to locate most of figurative language between these two extremes, with figures such as metaphor sitting roughly in the middle, adding and subtracting in equal measure from statement$_L$ to reach statement$_R$, allusions leaning more towards the left side (requiring more addition than subtraction) and hyperbolic expressions located further to the right (needing more subtraction than addition). I am keenly aware of the limits of understanding the reconstructive operations in this manner and I am far from advocating for a Lausbergian model of deviation categories for figurative language. I do however believe that this way of seeing can help to illustrate the different kinds of tools that are needed for a reasonable reconstruction under less than ideal conditions.

As mentioned above, these conditions require additional guidelines that justify a given reconstruction to a critical opponent. This is where the Gricean maxims and their Neo-Gricean counterparts can be a very helpful tool for reasonable reconstruction against resistance.[7] Their strength is perhaps most evident in the case of pure unexpressed premises. Take the following example:

P: I am certain Peter is quite pig-headed. He has been a teacher for more than twenty years.
A: I don't think that teachers are generally pig-headed.
P: I never said they were.

Leaving aside the more complicated question of the quantification of P's UP (i.e. are all teachers pig-headed or most, or the typical teacher?), it is clear that P's second turn violates the fifth commandment of the critical discussion by disowning P's UP. If requested to justify her reconstruction of P's UP against P's resistance, A

[7] Their practical usefulness is distinct from their theoretical status as ultimate reconstructive model. One thus does not need to take sides in the Gricean vs Neo-Gricean vs Post-Gricean vs Relevance Theory disputes to appreciate their utility for these purposes, but comp. Dynel 2018, 33ff. and Garmendia 2015, 40ff. for a defense of the Gricean group against some of the Relevance Theory challenges.

can point to Grice's maxim of relevance ("Be relevant", Grice 1989, 27). Assuming the absence of any reasons for violations, clashes or opting out, A may infer that P has flouted this maxim and reasonably heal the apparent violation by inserting P's UP above. Grice's maxim thus creates a prima facie case in favor of A's reconstruction which requires P to produce compelling reasons if he wants to eliminate the UP from his commitment set.

Levinson's principles can be of similar use for reconstructing UPs, as in the following example:

P: I actually enjoyed some of the past dinners we had together.
A: Really, only some. Which ones didn't you enjoy?
P: Oh, I didn't say that there were any I didn't enjoy.

While there is no clear flouting of Grice's relevance maxim in this short section without additional context, P clearly violates Levinson's Q-principle: "*Do not provide a statement that is informationally weaker than your knowledge of the world allows, unless providing an informationally stronger statement would contravene the I-principle. Specifically, select the informationally strongest paradigmatic alternate that is consistent with the facts*." (Levinson 2000, 76). A's recipient's corollary (ibid.) justifies her in presuming P's UP as "There were at least some dinners with you that I did not enjoy." Once again, P would need to provide independent reasons against this reconstruction if he wanted to disavow the UP in a reasonable manner.

While the practical utility of Grice's maxims and Levinson's principles is fairly evident for the more left, mostly addition-based, transfers, the case is more complicated for the middle and more right elements on the spectrum. Grice himself illustrates how his maxims can be used in the reconstruction of figurative language. His perhaps most famous examples in that regard are irony, metaphor, meiosis and hyperbole as flouting of the first maxim of quality ("Do not say what you believe to be false", Grice 1989, 27; Ibid. 34; comp. also Dynel 2018, 94ff.). Using this maxim for a justification of an interpretation against resistance leads to an evident problem at least in the case of irony.

The relevance maxim can produce a prima facie case in favor of an interpretation even against P's resistance because conversational relevance is a relatively objective standard.[8] "Believing something to be false" – as opposed to something actually being false – on the other hand, is a more subjective standard that gives P a considerable advantage. This is a minor problem for most figures (it is objectively highly unlikely that someone who calls Achill a "lion in the battlefield" believes that the Greek hero has four legs and is a feline, and someone

[8] This is not to say that it might not require interpretation, but instead that it is principally open and accessible to both parties.

who calls “all hands on deck” is almost certainly expecting the rest of the bodies to come along as well), but the case is more complicated for verbal irony. Grice’s own traitor example for irony works only because the contrast between statement_L and context is extraordinarily stark. Even for contrary irony this will not always be a reliable expectation, but in the case of Socratic irony it certainly becomes problematic.

4. RECONSTRUCTING SOCRATIC IRONY

Socratic irony is perhaps one of the clearest examples for a communicative phenomenon that is located on the far right of the interpretative spectrum, with the relationship of statement_L and statement_R being one of predominant subtraction and hardly any addition in some cases. In the first example introduced above, the content-based Socratic irony in Plato’s Gorgias, Socrates cannot reasonably be held to defend his claim_L that Callicles is an ideal model of knowledge, goodwill, and frankness, if challenged to do so by an antagonist (subtraction). He could however be reasonably expected to defend the claim that he thinks that Callicles thinks of Callicles as exemplary in these qualities (addition). He might also be reasonably held to his negative judgement of Callicles’ boisterous self-assessment.

In the second example, the case of clearly attributable content and lexicon-based Socratic irony, Tiffany Crutcher cannot be reasonably expected to actually believe that her deceased brother was a “big bad dude” (subtraction). If anything can be positively inferred from her statement_L at all, then it would be the evident fact that she is aware of the police recordings and their references to her brother, as well as her unsurprisingly negative assessment of the police statements (addition).

The final example provides a case of not clearly attributable, lexicon-based only, Socratic irony. Machine Head do not invoke a third party’s content, but instead only use the terminology of a vaguely defined group of racists and homophobes. The band cannot be reasonably expected to defend the racists and homophobe terminology (subtraction), and beyond a negative assessment of these groups, very little is reasonably added to their commitment set.

I hope that the above reconstructions of the three examples seem intuitively plausible to a benevolent reader – but they are lacking the kind of support against critical resistance that Grice and Levinson (and no doubt other similar models) can provide in the case of UP reconstruction. If an uncooperative antagonist would demand of Socrates to justify his emphatic praise of Callicles, or of Dr. Crutcher to explain her insulting treatment of her brother, or of Machine Head the racists and homophobe vocabulary in their song, how can they reasonably defend themselves and disavow their OPs?

Given that these right spectrum reconstructions are more likely to deal with potential fifth commandment violations of A, rather than P, the protagonist might have the advantage of a privileged insight into his own meaning. In the absence of resilient reconstruction guidelines this insight allows for a "That's not what I meant" at best – which is not a very satisfactory result. Can Grice and Levinson provide OP reconstruction guidelines similar to those we have for UPs?

The first maxim of quality is Grice's own example for contrary irony, but it doesn't quite work for Socratic irony. Sure, P might not fully (or at all) believe in the invoked content, but this lack of believe is only secondary to the principle positive believe in the truth of the implicit citation. P does believe that someone else holds what he says. The situation is even more complicated in the case of only lexicon-based Socratic irony, as here the content clearly does not have to be judged to be false, but rather the vocabulary to be inappropriate.

The maxim of relation and the maxims of quantity provide no evident starting point that could help with an OP reconstruction. The maxims of manner might occasionally help with the identification of supplemental irony markers especially in a literary context, where sudden obscurities or ambiguities may warn the reader of ironic subtext. However, the actual flouting that the author might commit is usually attached to stylistic effects that are parasitic or amplificatory to the actual irony and thus do not provide a reliable guideline either.

The situation isn't much more promising with Levinson's principles. His Q-principle and his I-principle have no discernible relationship to irony at all. His M-principle ("*Indicate an abnormal, nonstereotypical situation by using marked expressions that contrast with those you would use to describe the corresponding normal, stereotypical situation.*" Levinson 2000, 136) fulfills a similar role to that of Grice's maxims of manner. It can help to detect bonus markers of irony or amplificatory elements, but not the Socratic irony itself.

5. NO CONCLUSION

What remains is an aporetic conclusion. The Gricean and Neo-Gricean maxims and principles are very helpful for the reasonable reconstruction of unexpressed premises against critical resistance, but they seem less useful for their counterpart, the reasonable interpretation of overexpressed premises. A protagonist in a discussion under less than ideal circumstances, who uses of Socratic irony, is thus more vulnerable to (intentional or unintentional) fifth commandment violations of the antagonist than speakers who employ more easily decodable stylistic devices.

REFERENCES

Rhetorica ad Alexandrum (1957), trans. H. Rackham, Cambridge, MA: Harvard University Press.

Behler, E. (1998) Ironie. In. G. Ueding (Ed.) *Historisches Wörterbuch der Rhetorik* Bd, 4, 599-624. Tübingen: Max Niemeyer Verlag.

Booth, W. (1974) *A Rhetoric of Irony.* Chicago & London: University of Chicago Press.

Braungart, W. (2010). Eironeia urbana. In A. Burkhardt & B. Nerlin (Eds.) *Tropical Truth. The Epistemology of Metaphor and Other Tropes*, 323-338. Berlin: De Gruyter.

Burgers, C., van Mulken M. & Schellens P. (2011) Finding Irony: An Introduction6 of the Verbal Irony Procedure (VIP). *Metaphors and Symbol* 26 (3). 186-205.

Burgers, C., van Mulken M. & Schellens P. (2012) Verbal Irony: Differences in usage across written genres. *Journal of Language and Social Psychology* 31 (3). 290-310.

Burke, K. (1941). Four Master Tropes. *The Kenyon Review, 3(4),* 421-438.

Dynel, M. (2018). *Irony, Deception and Humour*. Boston & Berlin: De Gruyter Mouton.

Fogelin, R. (2011). *Figuratively Speaking*. Rev. Ed. Oxford: Oxford University Press.

Garmendia, J. (2015). A (Neo)Gricean Account of Irony: An Answer to Relevance Theory. *International Review of Pragmatics, 7(1),* 40–79.

Grice, P. (1989). Logic and Conversation. In P. Grice (Ed.), *Studies in the Way of Words* (pp. 22-40). Cambridge: Harvard University Press.

Haugh, M. (2013). Speaker meaning and accountability in interaction. *Journal of Pragmatics,* 48(1), 41-56.

Lausberg, H. (1949). Elemente der literarischen Rhetorik. München: Max Huber Verlag.

Lausberg, H. (1960). Handbuch der literarischen Rhetorik. München: Max Huber Verlag.

Levinson, S. (2000). Presumptive Meanings: The Theory of Generalized Conversational Implicature. Cambridge: MIT Press.

Muecke, D. C. (1969). *The Compass of Irony.* London & New York: Methuen.

Muecke, D. C. (1970). *Irony and the Ironic.* London & New York: Methuen.

Plato (1925) *Lysis. Symposium. Gorgias*. W. R. M. Lamb (Trans.). Cambridge, MA: Harvard University Press.

Quintilian, M. F. (2001). *The orator's education*. D. A. Russell (trans.). Cambridge, MA: Harvard University Press.

Saul, J. (2012). *Lying, Misleading and What is Said*. Oxford: Oxford University Press.

Simpson, P. (2011). "That's not ironic, that's just stupid" Towards an eclectic account of the discourse of irony. In M. Dynel (Ed.), *The pragmatics of humour across discourse domains* (pp. 33-50). Amsterdam: John Benjamins.

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

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition*. 2nd ed. Malden: Blackwell Publishing.

The Guardian (2016) *Video in Tulsa police shooting shows black man was unarmed with hands up*. Sep. 19, 2016. https://www.theguardian.com/us-news/2016/sep/19/tulsa-oklahoma-terence-crutcher-police-shooting. Retrieved on May 27, 2019.

van Eemeren, F. H. (2015) Argumentation Studies' Five Estates. In: Ibid. (Ed.) Reasonableness and Effectiveness in Argumentative Discourse. Cham et al.: Springer.

van Eemeren, F. H. (2018) *Argumentation Theory: A Pragma-Dialectical Perspective.* Cham: Springer Nature.

van Eemeren, F. H. & Grootendorst, R. (2004) *A Systematic Theory of Argumentation. The pragma-dialectical approach.* Cambridge: Cambridge University Press.

van Eemeren, F. H., Grootendorst, R., Jackson, S., & Jacobs, S. (1993). *Reconstructing Argumentative Discourse.* Tuscaloosa, AL: University of Alabama Press.

van Eemeren, F. H. & Snoeck Henkemans, A. F. (2017) *Argumentation. Analysis and evaluation.* 2. rev. ed. New York & London: Routledge.

Wilson, D., &, Sperber, D. (2012). Explaining Irony. In D. Wilson & D. Sperber (Eds.) *Meaning and Relevance* (pp. 123-145). Cambridge: Cambridge University Press.