



Meaning, Colouring, and Logic: Kaplan vs. Frege on Pejoratives

Significado, coloração e lógica: Kaplan vs. Frege acerca de pejorativos

 10.21680/1983-2109.2022v29n59ID29589

Ludovic Soutif

Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)

 0000-0002-9508-423X

lsoutif@gmail.com

Abstract: In this essay I consider Kaplan’s challenge to Frege’s so-called dictum: “Logic (and perhaps even truth) is immune to epithetical color”. I show that if it is to challenge anything, it rather challenges the view (attributable to Frege) that logic is immune to *pejorative* colour. This granted, I show that Kaplan’s inference-based challenge can be set even assuming that the pejorative doesn’t make any non-trivial truth-conditional (descriptive) contribution. Surprisingly, this goes against the general tendency to consider the truth-conditionally inert logically irrelevant. But I take it that Kaplan is right and take his examples to show that truth-conditional inertness need not entail inferential inertness. I end up assessing the Kaplan-Frege “debate” as giving edge to the former to the extent that clarity is achieved through Kaplanian inferences on what should be considered part of the *explanandum*.

Keywords: colouring, pejoratives, validity, Frege’s dictum, Kaplanian inferences

Resumo: Nesse ensaio contempla-se o desafio posto por Kaplan ao seguinte ditado atribuído a Frege: “Lógica (e, possivelmente, a própria verdade) é imune à cor epítética”. Eu argumento que, caso haja um desafio aqui, o que seria desafiada é a visão fregeana de que a lógica é imune à cor *pejorativa*. Dado isto, defendo que o desafio posto por Kaplan via inferências dedutivas se coloca ainda que se assumisse que o pejorativo não fizesse contribuição verocondicional (descritiva) não trivial. Surpreendentemente, isto vai de encontro à tendência geral em considerar o verocondicionalmente inerte logicamente irrelevante. Contudo, argumento que Kaplan tem razão e que seus exemplos sugerem que inércia verocondicional não necessariamente implica inércia inferencial. Finalizo mostrando que Kaplan leva vantagem no “debate” com Frege acerca de pejorativos em relação ao seguinte ponto: graças às inferências kaplanianas fica mais claro o que deve ser considerado parte do *explanandum*.

Palavras-chave: coloração, pejorativos, validade, ditado de Frege, inferências kaplanianas

Although Frege’s aim was not to provide a semantic theory for a natural language, he made, to say the least, valuable and enduring contributions to semantic theory, understood as the assignment of semantic values, and possibly, further meaning properties to natural language expressions. His arguing, notably

in Frege (1892), for the assignment to any well-formed linguistic expression of a sense (*Sinn*), in addition to its reference (*Bedeutung*) — if any—, is one — and, arguably, *the* most celebrated and disputed among his contributions. Another is the distinction drawn in his late writings (e.g. FREGE 1892a: 31, FREGE 1892b: 196 fn1, FREGE 1897: 151-4, FREGE 1906a: 209, FREGE 1906b: 214, FREGE 1906c: 102, FREGE 1918-19: 63-4) between the expression’s sense, that is, the thought (*Gedanke*) or thought-component expressed by it, on the one hand, and its colouring (*Färbung*), lightning or shading (*Beleuchtung*), on the other hand.¹ It matters to Frege to the extent that it is necessary, in his view, for *logical* purposes to separate the wheat (i.e. the truth-evaluable or conceptual part of a sentence’s content) from the chaff (its non-truth-evaluable or non-conceptual part, conveyed by some lexical item, grammatical construction or paralinguistic cue).² But it also matters to contemporary semantics, irrespective of Frege’s aims, insofar as an account is needed of those aspects of meaning that are responsible for conveying information beyond the sentence’s truth-evaluable content and to the extent that some instances of Fregean *Färbung* impart such information.³ Frege’s own account is often taken to be flawed on the grounds that it either misconstrues the phenomenon as purely subjective (DUMMETT 1981: 83-89; NEALE 1999: 38, 2001: 142; PICARDI 2006: 60; 2007: 500) or turns the retrieving of the non-truth-conditional information into a mere guesswork (KORTUM 2013: 2.2 & 2.11). However, both the subjectivist and the hint-based construals of Frege’s account can be shown to be off the mark if taken as purported general characterizations of the phenomenon of colouring (SANDER 2019).

In what follows, I shall argue that *if* there is anything flawed in Frege’s account the place to look at is his so-called dictum that the colouring, thought of in contemporary terms as an ingredient of conventional meaning, does not affect the thought expressed by the sentence in which the device occurs. For, if it can be shown, as Kaplan (1999) did show, that the position occupied by the colouring device —in Kaplan’s examples, pejorative epithets like “(that) bastard (PN)”, sentential interjections like “alas, *s*” or adverbs like “regrettably, *s*”, or ethnic slurs like “Chink” — in deductive inferences can make a difference to their validity, then there are objective reasons to doubt that the saying is true. And this is all the more puzzling that pejorative epithets or adjectives are often thought to make *no* truth-

¹ As Sander (2019: 380) rightly points out, Frege “was not particularly interested in establishing a consistent terminology”. Dummett (1981: 84) finds the *tone* metaphor more apt in English to translate the German terminology. But, even setting aside its prosodic connotations, the use of a single umbrella term in English tends to hide the diversity of phenomena gathered under it.

² See Frege (1906b: 214).

³ Since colouring is often involved in conveying something beyond the said, it is usually thought to be a pragmatic rather than semantic phenomenon (see e.g., KAPLAN 1999: 4 fn3). I concur, however, with Neale’s claim that, to the extent that colouring “concerns [in most cases] *the meaning of individual words*”, “it is not “a ‘merely pragmatic’ phenomenon” (1999: 45). Let me add that I also concur with Sander (2019: sec. IV & V) when he argues that, on Frege’s view, not all colourings, albeit semantically relevant, communicate some pragmatic content beyond the said. All in all, this is what turns this aspect of Frege’s contribution relevant to contemporary semantics.

conditional contribution, therefore, no contribution to the *thought* (to use Frege's terminology) otherwise expressed by the sentence in which they occur (GUTZMANN; GÄRTNER 2013: 4-5, GUTZMANN 2019: 14; *pace* see HAY 2013). So, how can this be? And how can Frege hold that the occurrence of an expression such as *cur*, which is a conceptual expression that seem to make some truth-conditional contribution, makes no logical difference whatsoever (all other things being equal)? This is, in a nutshell, the puzzle I shall be concerned with in this paper.

I used the conditional on purpose here, for it might be that Frege's account is, as a matter of fact, off the hook either because he relies upon a different taxonomy of non-truth-conditional (non-at-issue) contents not meant to carve "linguistic reality at its joints" (SANDER 2019: 389; see also 2021, forthcoming), or because his conception of validity is altogether different from the one used by Kaplan to challenge the dictum. So, I think we should keep an open mind on these issues and not necessarily take Kaplan's argument to be a knockdown one.

1. FREGE'S VARIANT OF THE DICTUM

Kaplan speaks of a *dictum* attributable to Frege (and Carnap) to the effect that "logic (and perhaps even truth) is immune to epithetical color" (1999: 4), while being fully aware that Frege (and, arguably, Carnap) never went so far as to take a stance, let alone make any official statement about the pejorative epithets' contribution (or absence thereof) to the thought expressed by the sentence in which they occur. In a footnote, Kaplan carefully modalizes his report as follows: "Epithets" Frege and Carnap *would have said* "do not contribute to 'cognitive content', and thus the study of their use belongs not to semantics but to pragmatics". (1999: 4 fn3, emphasis added)

Kaplan's line of reasoning seems to be roughly this: given what we know about Frege's notion of colouring, namely, that it is meant to cover all meaning difference that makes no difference to the sense (i.e., thought or "cognitive content") or truth-value (i.e., truth or falsity) of the sentence in which the corresponding device occurs, it is not hard to figure out what Frege *would have said* about the case of pejorative epithets *had he considered it*: being truth-conditionally inert, logic (i.e., deductive inferential relations among truth-conditional contents or thoughts, or the study thereof) remains unaffected by whatever difference they may otherwise be responsible for.⁴

⁴ I shall leave aside the question as to whether Kaplan is *right* in ascribing to Frege the view that since pejorative epithets are both truth-conditionally and inferentially inert the study of their use "belongs not to semantics but to pragmatics". Drawing the dividing line (in the case at hand) hinges, on the one hand, on how one takes colouring to be for Frege (namely, either as a purely pragmatic or a semantic/pragmatic phenomenon) and, on the other hand, on whether semantics is taken to be exhausted (on Frege's view) by the study of the sentences' truth-conditions.

Pejorative epithets are just one type of *c*(olouring)-device (SANDER 2019). Further *c*-devices include, for Frege, *inter alia* prosody, formal properties of speech and writing such as onomatopoeia, alliteration, assonance or rhyme, structural features such as the difference between the active and the passive voice, nominative-dative interchanges as in “M gave document A to N”, “Document A was given to N by M”, “N received document A from M”, words for logical particles such as “but” or “although” for the conjunction, adverbs like “still” and “yet”, sentence adverbs like “unfortunately, *s*”, interjections like “ah”, word-clusters featuring hypernyms (e.g. “walk”) and hyponyms (e.g. “stroll”, “saunter”), and more importantly for us here, pejorative common nouns such as “cur” from the “dog”-“cur” pair (see KORTUM 2013: Part II for a comprehensive review). If Frege never went so far as to take a stance on the nature of the contribution of pejorative epithets to the thought expressed by the sentences in which they occur and its consequence for logic, he went so far as to take one on the issue of how many thoughts are expressed by a sentence featuring the pejorative term “cur”. This surely matters, since if the answer is that a thought is expressed by it in addition to the one expressed by its neutral sentential counterpart (all other things being equal), to the extent that “thought” designates for Frege that part of the sentence’s content responsible for its inferential potential and “logic” the study of inferential relations among thoughts, the fact that two thoughts instead of one are explicitly expressed in this case may have bearings upon the question as to whether logic is (or is not) immune to pejorative colouring. So, it is worth delineating Frege’s argument, not least because it provides us with less conjectural a way to make sense of the so-called dictum, or rather, some suitable variant of it.

1.1. Frege’s Argument Against the Many-a-Thought Objection

Consider one of Frege’s favourite examples (see FREGE 1897: 152ff.; 1979: 140ff.), namely the sentence pair below featuring a *c*-device (the pejorative “cur”) and its neutral counterpart (the non-pejorative “dog”).

- (1) This dog howled the whole night.
- (2) This cur howled the whole night.

On Frege’s view, (1) and (2) are — to use non-Fregean terminology — extensionally as well as intensionally equivalent, which means that replacing “dog” by “cur” in (2) — all other things being equal — does not alter (1)’s truth-value: if (1) is true, so is (2)⁵; nor does it alter its sense (*Sinn*), since the thought (*Gedanke*) — Frege’s other word for the *sense* of an assertoric sentence — in (1) and (2) “is the same” (1897: 152). It does not follow, however, that “dog” and “cur” are, for him, synonyms.⁶ For them to be synonyms (i.e., substitutable in all contexts *salve*

⁵ “[...] the use of the word ‘cur’ does not prevent us from holding that the second sentence is true as well.” (FREGE 1897: 241; 1979: 140).

⁶ Here I disagree with Picardi (2006: 62) and side, rather, with Kortum (2013: 102) and Horn (2013: 157 fn10).

significatione), no meaning difference ought to be allowed to hold between them. But “it appears to be Frege’s view”, as Neale (2001: 140) rightly points out, “that sameness of meaning is guaranteed by neither sameness of reference nor sameness of sense”, “in some intuitive sense of our word ‘meaning’”. This is, I take it, what makes the example so telling here: sameness of both reference (truth-value) and sense (thought) is no guarantee that no meaning difference holds between (1) and (2). To be fair, Frege (1897: 152-3) does not speak of a *meaning* difference holding between “dog” and “cur” — by compositionality, between (1) and (2). It is in these terms, nevertheless, that the difference is usually couched in the contemporary literature, a difference that Frege’s colouring (*Färbung*) and cognate notions are meant to capture. What Frege *does* say in the passage from *Logic* alluded to above is that, despite their expressing the same thought, (1) and (2) differ due to the presence of a non-neutral term in (2) and the absence thereof in (1). So, the question to ask now is: how are we to understand the non-neutrality of “cur”?

The English translation suggests that it is a matter of psychological association — on a subjectivist (Dummettian) construal of the notion of colouring —: the word “cur”, the translation reads, “certainly has unpleasant rather than pleasant associations and puts us rather in mind of [*sic*] a dog with a somewhat unkempt appearance” (1979: 140). Since, the argument runs, different subjects can associate different subjective representations (or attitudes) with the word “cur”, it seems to follow that the meaning difference at stake here cannot be accounted for as it should be, namely, as an objective and conventional one (DUMMETT 1981: 85; see also DI FRANCO 2014: 2a for the particular case of the word “cur” *qua* c-device). This construal, however, is not backed by the original German wherein the difference between the attitudinally neutral “dog” and non-neutral “cur” is presented, rather, as a matter of hint (*Wink*) given by the latter as to how the dog is to be pictured, namely, as “somewhat rough-looking”.⁷ Assuming the difference at stake is a meaning difference, this seems to point toward a semantic/pragmatic interpretation of colouring on which it is essentially a matter of F-implicature — a forerunner to Grice’s conventional implicature (HORN 2007, 2013).

Whatever construal is most faithful to Frege’s original view — I am prone to think that the latter is —, the point is that the difference in both cases is believed by Frege to be irrelevant for logic, since logic is only concerned, in his view, with that which is *part* of the thought (*Gedanke*) expressed in the sentence and whatever is expressed in (2) by the word “cur” itself — typically, for Frege, an

⁷ The German reads: “(...) einen Wink, sich den Hund etwas ruppig vorzustellen.” (FREGE 1897: 152)

attitude of disdain on the speaker's part⁸ — is *not* part of it, but only attached to it by convention.⁹

Frege rarely makes a case (even indirectly¹⁰) for the view that whatever is conveyed by c-devices has no relevance for logic and, hence, leaves logic unaffected. So, it is fortunate that in the case, at least, of pejorative nouns an argument *is* provided by him.

To begin with, Frege considers the competing view according to which more than one thought is expressed and asserted by (2) — when used with assertoric force. More specifically, the competing view holds that two thoughts, instead of one, are expressed and asserted by (2): the thought that this dog howled the whole night (i.e., the one expressed by (1)) and, in addition, the thought that the speaker has a poor opinion of the dog (call them, respectively, *T1* and *T2*). Note that it is not too far-fetched a view. Assuming the word “cur” encodes as a matter of conventional meaning some additional conceptual material featuring the property of being worthy of poor opinion (for displaying such and such appearance), a way to explain the semantic difference between (1) and (2) would be to do so in terms of a conjunction of thoughts (or a more complex thought) expressed and asserted by (2), as opposed to a single (or more simple) thought expressed and asserted by (1) — when used with assertoric force. It is true that a gap remains between the thought that the speaker has a poor opinion of the dog and the thought that the dog is worthy of poor opinion. But the gap can be bridged by arguing that the latter (objectively) grounds the former and that the former is entertained by the speaker in answer to the latter.¹¹

However likely the competing view may be, Frege (1897: 151-2) considers it as a potential objection to his own identity view — the view that just *one* thought is expressed or asserted by (2), the one also expressed or asserted by (1) — and refutes it by *modus tollens*.

The argument runs basically as follows. Let us call the competing view *MTV* (for Many-a-Though View).

⁸ Again, the English translation is below the point. The original German reads: “Wer ihn ausspricht, äußert damit allerdings eine gewisse Geringschätzung; aber diese gehört nicht zum ausgedrückten Gedanken.” (1897: 152)

⁹ On “the importance for logic” of “the distinction between what is part of the thought expressed in a sentence and what only gets attached to it [the thought]”, see FREGE (1897: 153). See also FREGE (1892b: 196 fnG). Since the attitude of disdain, on Frege's view, gets only attached *by convention* to the thought expressed by (2) and conventions (i.e., tacit agreements among users) can change over time, it may *come to be* part of the thought expressed in the sentence and hence relevant for logic. But, as Frege himself points out (1897: 153), the distinction is not thereby “obliterated”. It is just that more “things” become part of the thought and hence turn themselves logically relevant.

¹⁰ As Sander (2019: 382) rightly points out, that c-devices are logically irrelevant (or inferentially inert) is not something that “Frege explicitly says”, but rather a “corollary” from their truth-conditional inertness.

¹¹ See Bach (2018) for a contemporary proposal along these lines concerning slurs — further members of the class of derogatives.

P1. If MTV is right, then if T2 is false, then the thought expressed by (2) is false (being a *conjunction* of T1 and T2)

P2. (2) is in the scenario isn't false (it is possible, at least, to withstand the temptation to deem it so)

Conclusion. MTV isn't right.

The crucial premise, of course, is P2. Frege vindicates its truth by resorting to the distinction between expressed and unexpressed (yet induced) thoughts. Consider one of his examples of deceiving behaviour: the case of a person *pretending* she is sad using a sad tone of voice. Frege (correctly) argues that the false impression created by the using of a c-device (in the case at hand, of the prosodic variety) is compatible with the claim that the thought expressed, as opposed to the one induced, is *true*. If this is right, assuming a parallel case can be made for (2), what we get is, arguably, a refutation of MTV.

1.2. Troubles for Frege's View

Although the variant of the so-called dictum ("logic is immune to pejorative colour") it is meant to support relies upon a contentious assumption (to be reviewed in section 2.2 below), I take Frege's argument against MTV to go through.¹² The trouble lies, rather, in his taking for granted that "dog" and "cur" — and, by compositionality, (1) and (2) — are extensionally and intensionally equivalent.

For one thing, it does not seem reasonable to hold (unless for argument's sake) that "dog" and "cur" are substitutable *salva veritate* in *all* extensional contexts of utterance. Suppose my neighbour who owns the dog utters (3) in response to my uttering (2) with assertoric force.

(3) That dog is not a cur.

If Frege is right (i.e., if "dog" is replaceable by "cur" in all extensional contexts without altering the sentence's truth-value), my neighbour is making a false statement, since "that dog is not a dog", when asserted, states a contradiction. But plainly no false statement is being made here, let alone a necessarily false one.¹³ The same holds for a bunch of quantificational statements such as (4) and (5) below: if Frege is right, then (setting aside the possibility of reference failure) they express necessary (provided meaning postulates are laid down, even *analytic*) truths, which is plainly *not* the case; in contradistinction, say, to (6).

¹² I do not think, for instance, that an argument against the truth of the second premise drawing on something like Bach's speech-report argument (in BACH 2018) against hybrid expressivism would be compelling. But delineating a counterargument is beyond the scope of this paper.

¹³ Here I am drawing on Picardi (2006: 62).

- (4) All dogs are curs.
- (5) Some dogs are curs.
- (6) All curs are dogs.

For another thing, contrary to what Frege assumes, (1) and (2) are not equipollent sentences in Frege's sense (see 1906b: 213; 179: 197), since it is clearly possible to recognize the content of one as true without recognizing straightaway the content of the other as true. This means that in case equipollence is picked as an identity criterion for thoughts, the thought expressed (and asserted) by (1) is not the same as the thought expressed (and asserted) by (2) and *vice versa*. Or if they are, it is only on the assumption that the difference is truth-conditionally inert and hence logically irrelevant, which is precisely what the argument was meant to *establish*.

1.3. Reply (on Frege's Behalf) and a Rejoinder

An obvious reply on Frege's behalf is that he could not care less about giving an account of the exact nature of the pejorative's contribution to that which is conveyed by utterances of the corresponding pejorative sentences (in contrast to their neutral counterparts) and that his terminological sloppiness regarding all that is not part of the thought (assertible content) explicitly expressed in a sentence mirrors just this fact. He could not care less, the reply goes on, because all he is interested in is separating the wheat (*i.e.*, the part of the sentence-content relevant to the study of truth and truth-preserving inferences) from the chaff (*i.e.*, the part that isn't, *although it may seem to be*). Carrying out this critical task is what really matters to Frege — one might add — insofar as that which is relevant for logic (for Frege, thoughts or assertible contents and their inferential relations) always comes mixed up in human reasoning with features that are not and that he gathers under the heading of *colouring*, *shading* or *scent* — however heterogenous these features may turn out to be. To that end, natural languages are more of an obstacle than a help, not least because they present us with more (meaning) differences than really matter to logic. Sentences (2)-(6) are — one might conclude — good cases in point.

I do not deny, of course, that this is how Frege's 'thought'- and 'colouring'-talk ought to be understood, namely, as having as its main goal not to deliver the right (or best possible) account of the pejoratives' semantic/pragmatic contribution *qua* c-devices, but to isolate the logical from the non-logical by means of a criterion (whatever makes any difference to the thought expressed in a sentence and its relations to other thoughts is worth being considered as logical). Still, if Frege has a point in arguing in favour of the truth-conditional inertness of the "thought" conveyed by the pejorative noun, the semantic argument (delineated in 1.1 above) can be taken to vindicate, if only indirectly, some variant of the so-called Fregean

dictum, namely, the saying that *logic is immune to pejorative colour*. It is time now to assess the saying (regardless of whether Frege explicitly endorsed it), which leads me to consider Kaplan's challenge to it.

2. KAPLAN'S CHALLENGE

As is well known and as was already noted, Kaplan (1999) sets forth examples of deductive inferences featuring pejorative terms (typically, epithets, but also possibly sentential interjections and adverbs, or ethnic slurs) the validity of which turns on the position occupied by the colouring device within the argument. In Kaplan's view, they are meant to challenge more or less straightforwardly Frege's (and Carnap's) so-called dictum: "Logic (and perhaps even truth) is immune to epithetical color" — in fact, we saw that if they are to challenge anything, it is rather the dictum's variant: "logic (and perhaps even truth) is immune to pejorative colour". In section 2.1 below I review Kaplan's argument. Since there is more to it than meets the eye, a bit of a reconstruction is needed. I go on defusing a possible objection on Frege's behalf relying upon the assumption that in order to affect logic (i.e., the argument's validity) the content semantically encoded by the pejorative must be part of the thought (in contemporary terms, of *what is said*). This is, as Kaplan (1999: p. 4 fn5) himself points out, an assumption commonly made in semantics and pragmatics and Frege is surely no exception. However, following Kaplan, I shall argue that this is an assumption one need not grant, and that Kaplan's challenge can be set even assuming that the expressive content encoded by the colouring device is truth-conditionally inert while conveying a derogatory attitude along another semantic/pragmatic dimension (sec. 2.2).

2.1. Kaplan's Argument Reconstructed

Kaplan's argument can be reconstructed as follows: if the dictum is true, then the semantic content encoded by the pejorative term (*qua c-device*) does not affect logic (i.e., logic is immune to it); it does; therefore, the dictum isn't true (i.e., logic *isn't* immune to pejorative colour). The key premise is, of course, the second, namely, the premise to the effect that the semantic content encoded by the pejorative term (formally represented by Kaplan as the set of contexts in which is the term is correctly used) does affect the argument's validity. What speaks in favour of its truth is that the rules for their correct use produce, as Kaplan points out, "a distinctive and deviant pattern of logical consequence" (2004: [18:07]).

Consider, for instance, the rule for the correct use of the pejorative epithet "damn" in "that damn Kaplan". It is correctly used just in case the agent of the context (the speaker) holds a derogatory attitude towards the expressive's target at the time and the world of the context of utterance and that is *all* the pejorative

is semantically contributing to the sentence as a whole. Since, by hypothesis, no descriptive (truth-conditional) contribution is made by the pejorative term itself¹⁴, its occurring in the argument should not alter the argument's validity —the underlying assumption being that the relation of logical consequence between the premises and the conclusion can only be altered if the term makes a non-trivial descriptive or truth-conditional contribution (see next section). However, consider arguments A1 and A2 or, alternatively, A1* and A2* below. The Fregean prediction¹⁵ is plainly incorrect, since the argument's logical status *is* altered by the occurrence of the epithet in the conclusion: intuitively, A1 and A1* are valid, while A2 and A2* are not. This means that, while truth-conditionally inert, the c-device is *not, pace Frege, inferentially so*.

(A1)

That damn Kaplan was promoted; therefore, Kaplan was promoted.

(A2)

Kaplan was promoted; therefore, that damn Kaplan was promoted.

(A1*)

Alas, Kaplan was promoted; therefore, Kaplan was promoted.

(A2*)

Kaplan was promoted; therefore, Kaplan was, alas, promoted.

If the second premise is indeed true, then the argument goes through, thereby providing objective reasons to doubt that the variant of Frege's saying is true.

2.2. Defusing a Fregean Objection

It might be objected that Kaplan's inference-based challenge is misguided because whatever is conveyed by the pejorative (for Frege and Kaplan, the speaker's derogatory attitude) does not affect the sentence's truth-conditions (the "thought" expressed by it) and hence has no inferential potential. Being truth-conditionally equivalent to its neutral sentential counterpart, the pejorative sentence has no more inferential potential than the latter. Or if it does, as in Kaplan's inferences, it is because the expression's inferential potential was misplaced. This is the kind of objection one can easily imagine giving on Frege's behalf. Sander voices it as follows:

¹⁴ "Assume that the epithet "damn" in "That damn Kaplan" expresses derogation *and has no descriptive content* — allow me this in order to make my point." (KAPLAN 1999: 5. Emphasis added).

¹⁵ Remember that for Frege the pejorative noun does not contribute to the thought expressed by the sentence in which it occurs. It is descriptively (i.e., truth-conditionally) and hence *inferentially inert*. So, although Frege himself made no such prediction, it can be made on his behalf.

(...) If Frege had known Kaplan's (...) claim that the sentence 'Kaplan was promoted' follows from 'That damn Kaplan was promoted' but not vice versa, he would have regarded this as seriously misguided. For Frege, by hearing a sentence that contains words like 'damn' or 'cur' you do not learn (*erfahren*) more than by hearing their neutral counterparts, though he would concede that a speaker who uses such words conveys (*aüssern*) some kind of contempt (*Geringschätzung*) or the like. However, what a speaker conveys by some word does not necessarily affect the inferential relations between sentences that contain them. (SANDER 2019: 382)

The objection relies upon the assumption that only what is truth-conditionally relevant (*what is said*, in Grice-Kaplan's vocabulary) can affect logic (i.e., relations of logical consequence between thoughts). This is, as Kaplan (1999: 4 fn5) himself points out, a "common assumption" to make made not only by Frege, but by most contemporary authors concerned with the semantics-pragmatics divide, starting with Grice. But there are reasons to believe that, while common, it is unwarranted and that Kaplan's inferences are meant to show, precisely, that while there is a sense in which A1 (*) and A2 (*) can be said to be equally valid, there also is a sense in which they can be said not to be, namely, the sense of "validity" in which expressive elements such as the speaker's (derogatory) attitude are allowed into the picture. If, the counterargument runs, expressive elements such as these had no inferential potential, it would make little sense to speak of a difference in logical status between A1 (*) and A2 (*); it makes sense; *ergo*, they do have inferential potential.

Note that this is perfectly compatible with the claim that the pejorative *qua* c-device does not contribute to what is said (in Frege's terminology, to the *thought* expressed), so that Kaplan's challenge can be set even assuming that no truth-conditional contribution is being made *or* that a contribution is made by the pejorative along another semantic/pragmatic dimension than the truth-conditional one. Kaplan makes clear that this holds for Grice's analysis of the meaning of words for logical particles (further members of the class of c-devices) in terms of conventional implicature. But it is not too much of a stretch to say that the same holds for Frege's "analysis" of the meaning of pejorative nouns in terms of (unasserted) hints given by the speaker, considering that F(rege)-implicature is rightly taken by some authors (in particular, Horn) to be a precursor to Grice's notion of conventional implicature:

The arguments I will present are meant to show that even accepting Grice's analysis, the logic is affected by the choice of particle, as it should be on my view of logical validity as the preservation of truth-plus rather than (merely) descriptive truth. If this is correct, then generations of logic teachers, including myself, have been misleading the youth. Grice sides with the logic teachers, and though he regards the expressive content as *conventional* and hence (I would say) semantic (...), he categorizes it with the maxim-generated *implicatures*. (KAPLAN 1999: 20-21; apud HORN 2013: 150).

If Kaplan is right, that is, if the challenge can be set even assuming that the c-device makes no truth-conditional (descriptive) contribution because truth-conditional inertness need not entail inferential inertness, then not only will it not

do to reply on Frege's behalf that offloading the c-device's inferential potential to another dimension of meaning is "misguided" since the latter is logically irrelevant, but a natural way to interpret the challenge is as casting doubts over the truth of the variant of the dictum attributable to Frege.

I am already here in the business of assessing the "debate" between Kaplan and Frege over the semantics and logic of pejoratives, assuming the commensurability of their respective contributions. But a proper assessment needs to take a step back and consider whether there is any "debate" at all.¹⁶

3. ASSESSING THE "DEBATE"

3.1. Do They Talk Past Each Other?

One might indeed argue that there is no debate at all and that Frege and Kaplan just talk past each other. For one thing, Frege's taxonomy of non-truth-conditional (not-at-issue) contents is a logic-driven, purpose-related one with no overlap with current taxonomies, whereas Kaplan's is an attempt to capture some features of expressive content that partially overlaps with current taxonomies.¹⁷ For another, the relevant notion of (logical) validity for Frege is, presumably, the classical one defined over truth — a deductive argument is valid, on Frege's view, just in case it is truth-preserving¹⁸ — while Kaplan (1999) redefines (logical) validity either as information delimitation, or as truth-plus preservation. Let me elaborate a bit.

Frege's taxonomy of colourings, or rather, what may be reconstructed as such (see SANDER 2019: 388; 2021, forthcoming) was devised primarily for the purpose of separating what is relevant to a study of inferential relations between thoughts (recognized as true) from what is not yet comes mixed up with it in ordinary language — for *logical* purposes, that is. Since what determines the importance of a topic is ultimately the theoretician's purpose, what is considered essential, for instance, by literary critics, linguists, or meta-ethicists, may not be so considered by the logician and vice versa. As a result, distinctions that may seem essential to the former may be overlooked by the latter (and vice versa). Sander quotes a passage from *The Thought* wherein Frege makes this point quite clear. After giving examples of distinctions that, from the logician's point of view, "make no difference to the thought" (typically, distinctions involving the use of c-devices

¹⁶ Here and in what follows, I shall use scare quotes around the word *debate* for, historically speaking, it is one-sided.

¹⁷ Unsurprisingly, since Kaplan (1999) helped shape them.

¹⁸ I am not using the biconditional and do not take "just in case" to stand for it, for being truth-preserving is not sufficient, in Frege's view, for validity. In addition, the thought expressed in the conclusion must be *recognized as true* in virtue of having been inferred from thoughts *recognized as true* in the premises. See CURRIE (1987).

such as words for logical connectives, adverbs, the use of the passive vs. active voice and nominative-dative interchanges), he goes on noticing:

It is just as important to ignore distinctions that do not touch the heart of the matter, as to make distinctions which concern what is essential (*das Wesentliche*). But what is essential depends on one's purpose (*Zweck*). To a mind concerned with the beauties of language, what is indifferent to the logician may seem to be just what is important. (FREGE, 1918-19: 64; Sander's modified translation in SANDER 2019: 389)

A good case in point is the distinction between words for logical connectives such as “but” or “although” and pejorative nouns such as “cur”. Both are classifiable in Fregean terms as c-devices. Still, the emphasis in current taxonomies of non-truth-conditional contents is on noteworthy behavioural differences between them when embedded in doxastic operators. Consider belief reports such as (7) and (8) below, uttered by A:

(7) B believes that C is poor *but* honest.

(8) B believes that *that cur* howled the whole night.¹⁹

As Sander rightly points out, assuming their semantic/pragmatic contribution is not exhausted by the meaning features responsible for their truth-conditional profile, there is a striking difference here in projection behaviour between the (italicized) c-devices evidenced by the fact that (8) allows two readings on which the attitude of disdain expressed by the pejorative noun either is A's or attributable to B (when uttered by A), while (7) allows just one reading on which A takes B to be committed to some sort of contrast between (C's) poverty and honesty. In (7) the doxastic operator acts as a “plug” (KARTTUNEN 1973), that is, a predicate that blocks off the commitment of the embedded clause, while in (8) it acts as a “hole” (ibid.), that is, a predicate that (on one possible reading, at least) passes it up the tree.

On Frege's view, this distinction, grounded in a behavioural difference in belief reports, can be ignored insofar as it is (allegedly) irrelevant to logic. What matters to the logician (as opposed, say, to the literary critic) is that both devices be truth-conditionally and hence (for Frege) inferentially inert. And they are, on Frege's view, because whatever content is conveyed by the device isn't expressed (or asserted) by the sentence, but only induced by way of hints. By contrast, Kaplan takes it not only that unpluggability (scopelessness, non-displaceability) need be accounted for as a distinctive feature of expressive (derogatory) content, but that it has bearings upon logic in the following sense: some arguments look valid only if we overlook the fact that some derogatory content is expressed the hallmark of which is to be scopeless (unpluggable, non-displaceable).²⁰ Conversely, taking this

¹⁹ The (slightly modified) examples are Sander's in SANDER (2019: 388). Hereafter, I am drawing heavily on his analysis. My emphases.

²⁰ Although Kaplan (1999) does not use the current terminology, he tackles the issue under the general heading of scope. On scopelessness (unpluggability, non-displaceability, hyperprojectability) as a distinctive and,

fact into account leads us to consider a notion of (logical) validity significantly different from the classical (Fregean) notion either because it is not defined over truth or because, in case it is, truth is not enough to capture it. This corresponds to Kaplan's proposal to redefine logical validity either in terms of information delimitation (containment) or truth-plus preservation.

3.2. Kaplan's Edge

Nevertheless, I take it that the "debate" still stands, and that Kaplan has the edge over Frege. It stands because whether a deductive inference is valid is not a matter of decision on the logician's part. Kaplan cashes this out in terms of logic (or logical theory) not being a "purely stipulative science" (1999: 4 fn4), meaning that we must rely up to a certain point "on our pre-theoretical intuitions about logical consequence" (ibid.). Indeed, the validity of $A1(*)$, as opposed to the non-validity of $A2(*)$, can be considered a *datum* to be accounted for by any logical theory no matter how the notion of logical consequence is defined.²¹ So, even though Kaplan and Frege can be viewed as talking past each other on account of working with different definitions of logical consequence, the "debate" is still on because our pre-theoretical intuitions present us with a difference in logical status that must be acknowledged prior to any explanatory task. And it will not do to argue along Fregean lines that this is confused because the colouring (or shading) is inessential to the logician's purpose, since whether the behaviour of non-truth-conditional content is (or is not) relevant to the validity of arguments is not "a purely stipulative" matter. The mere fact that clarity is achieved here regarding the explanandum-explanans distinction gives Kaplan a certain edge. What seems to speak also in favour of Kaplan is that his logical theory is, arguably, equipped to account for a wider range of valid deductive inferences than Frege's, as the examples of deductive inference involving pejorative terms suggest. This, however, need be verified rather than taken for granted.

3.3. A Risky Bet

Consider one of the options envisioned by Kaplan, namely the redefinition of logical validity as information delimitation (contention). The hallmark of this strategy is to avoid relying on the notion of truth, given the variability and shakiness of our intuitions about its applicability in the cases under scrutiny, relying instead on both the notion of semantic information (carried by a linguistic expression in virtue of its conventional meaning) and "our more stable" and "widely shared" intuitions about logical consequence (KAPLAN 2004: [33:30]).

arguably, defining feature of expressive (derogatory) content, see amongst others POTTS (2007), SAUERLAND (2007), GUTZMANN (2019), SOUTIF; MÁRQUEZ (2020).

²¹ Note it has been considered so in the literature on pejorative content. See e.g. HOM (2010), DIFRANCO (2014).

Information delimitation (contention). There must be no semantic information in the conclusion that is not already contained in the premises.

This surely is a way to explain the datum, since the difference in logical status can be explained by the fact that A2(*), in contrast to A1(*), violates the constraint: although the conclusion does not *say* more than is already said in the premise, it contains more *semantic information*, namely the information that the speaker despises (or holds whatever negative attitude towards) the expressive's target (suits to the use of "damn") or, as far as A2* is concerned, laments the fact that Kaplan was promoted, in addition to the truth-conditional information (that Kaplan was promoted) shared by the premise. By contrast, the datum is not (and cannot) be explained using the classical notion of logical validity as truth-preservation, since A1(*) and A2(*) both are truth-preserving. So, Kaplan's (first) attempt to redefine logical validity gives him *prima facie* an edge (over Frege) to the extent that he is thereby able to account for a wider range of data. Still, this comes at a cost, as the strategy faces two unresolved issues.

First, the constraint is loose (or insufficiently tight). Consider a deductive inference such as $\neg A$; therefore, A (VALTONEN 2019: 7). The constraint is arguably met: the conclusion does not contain more information than the premise. Yet, the former does not seem to logically follow from the latter. Second, defining logical validity as information contention turns the validity claim into an analytic truth, given that deduction can also be defined in these terms. This does not look right: validity must be a "synthetic" property of deductive inferences if we are to allow the possibility of invalid deductive inferences.

Now consider Kaplan's other option, namely redefining logical validity in terms of a broader notion of truth-plus (see 1999: 6-8). A pejorative sentence is true, on this notion, only if it is descriptively *and* expressively correct (*i.e.*, only if what it says is the case and the attitude purportedly expressed by the pejorative is held by the speaker). Take, for instance, "that damn Kaplan was promoted". It is true-plus only if Kaplan was promoted, and the speaker despises Kaplan. This granted, a broader notion of validity-plus can be defined as follows:

Validity-plus. A deductive inference is valid-plus only if it is truth-plus-preserving (*i.e.*, truth- plus attitude-preserving).

Again, an important merit of this attempt at redefining (logical) validity is that it explains the datum: on the new definition, A1(*) is while A2(*) is not truth-plus preserving (hence, is invalid). And this seems to give Kaplan an edge over Frege for the reasons set out above. This definition is meant to *improve* on the notion of validity classically defined as truth-preservation. This means that, given the definition, all truth-plus-preserving deductive inferences are expected to be truth-preserving, but not vice-versa. And this is, indeed, the case with A1(*) and

A2(*): the former is truth-plus and truth-preserving, therefore valid (both in the classical and the extended sense of the word “valid”) while the latter is truth-preserving (*i.e.*, classically valid) albeit invalid in the extended (truth-plus-preserving) sense of the word.

The trouble is that some deductive inferences that are presumably truth-plus-preserving *aren't* truth-preserving, thereby unfulfilling the expectation. Consider the following couple of arguments:

(A3) Fido is a dog and I regard dogs as contemptible; therefore, Fido is a cur.

(A4) Fido is a cur; therefore, Fido is a dog and I regard dogs as contemptible.²²

Let us assume for argument's sake that “I regard dogs as contemptible” is the correct sentential paraphrase for “cur” and that the expression is correctly used only if the paraphrase is true at the world and the time of the context of utterance — this is in line with Kaplan's own analysis. Intuitively, both arguments are, in contrast to A1 and A2, truth-plus-preserving since the descriptive content *and* the attitude are preserved in the conclusion. Still, A4, as opposed to A3, is *not* truth-preserving, since there is at least one interpretation on which the conclusion is false and the premise true: the one on which the pejorative is *not* used as it ought to be, namely as expressing or displaying a derogatory attitude on the speaker's part.²³ If this is right, then *validity-plus* does not do what it was meant to, namely, improve on the classical definition of (logical) validity.

Either way, what initially seemed to give Kaplan the edge over Frege eventually turns into a risky bet on the prospects of a redefinition of the notion of logical validity.

4. CONCLUSION

Let me sum up the results achieved in this essay on the Kaplan-Frege “debate” over pejoratives. To begin with, I argued that although Kaplan conjecturally ascribes to Frege a dictum that concerns the consequences for logic of the truth-conditional inertness of the meaning of pejorative *epithets*, there are solid grounds for the ascription to Frege of a variant of the dictum — something like: “logic (and perhaps even truth) is immune to *pejorative* colour”. The grounds are Frege's argument against MTV to the extent that it concerns the logical bearings of the occurrence of pejorative nouns (such as “cur”) in sentences. I

²² I am adapting an example from Valtonen (2019). Note that the example does not work with pejorative epithets, presumably because they have no (detachable) neutral counterpart. Since Frege's pejorative nouns do have such counterparts, it can be easily adjusted using one of Frege's favourite examples of c-device.

²³ Note that, in this case, while not *true-plus* the premise is still *true* because it is enough that it be truth-conditionally equivalent to “Fido is a dog”.

showed that, despite the difficulties faced by Frege's identity view, the argument goes through yet relies on the contentious assumption that the relation of logical consequence between the premise(s) and the conclusion can only be affected provided the term makes a non-trivial descriptive (truth-conditional) contribution to the thought expressed (or asserted). As I understand them, Kaplan's examples of difference in logical status between deductive inferences depending on the whereabouts of the pejorative within them challenge just that. This is shown by the fact that the challenge can be set even assuming (as Frege has it) that the pejorative is truth-conditionally inert. The proper way to understand Kaplan's challenge is, in my view, to understand it as challenging the Fregean idea that the pejorative epithet, being truth-conditionally inert, has *no* inferential potential. I take it that Kaplan is right in arguing against Frege that to the extent that *there is* a difference in logical status between A1 and A2 (the former being valid, the latter invalid), the pejorative epithet *has* inferential potential depending on where it occurs in the inference. And like Kaplan, I take it, *pace* Frege, that this is a *datum* to be accounted for by a logical theory. Kaplan's theory seems equipped to account for a wider range of deductive inferences, including expressively valid (as opposed to invalid) ones, and this gives him *prima facie* an edge over Frege in the "debate". However, I ended up showing that Kaplan's attempts to redefine the notion of logical validity (either as information delimitation or validity-plus) in order to account for the datum turns into a risky bet on the prospects of such redefinitions. If (as I am inclined to believe) the prospects aren't good, then redefining logical validity either way gives Kaplan no real edge in the "debate". What gives him an edge, though, is to have achieved clarity on the issue as to whether the pejorative's inferential potential is part of the explanandum.

ACKNOWLEDGMENTS

Research for this paper has benefited from the support of the CNPq (*Conselho Nacional de Desenvolvimento Científico e Tecnológico*, Brasil) via a Senior Post-Doctoral Grant (#101052/2020-0) and a Research Productivity Grant (#305436/2020-1). It has also benefited from the support of the CAPES (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*, Brasil, #001). I am grateful to the participants of the Symposium on Expressivity held in May 2022 at the Universidad de la República Uruguay (Montevideo) for challenging discussions and helpful suggestions.

REFERENCES

AUXIER, R.E.; HAHN, L. E. (eds.). *The Philosophy of Michael Dummett*. Chicago, La Salle, ILL.: Open Court, 2007.

- BACH, K. *Loaded Words: On the Semantics and Pragmatics of Slurs*. In: D. SOSA (ed.) (2018), pp. 60-76.
- COSENZA, G. (ed.) *Paul Grice's Heritage*. Turnhout: Brepols, 2001.
- CURRIE, G. Remarks on Frege's Conception of Inference. *Notre Dame Journal of Formal Logic*, 28 (1), pp. 55-68, 1987.
- DI FRANCO, R. Pejorative Language. *Internet Encyclopedia of Philosophy*, 2014.
- DUMMETT, M. *Frege: Philosophy of Language*. Cambridge, MA: Harvard University Press, 1981.
- FREGE, G. Über Begriff und Gegenstand. *Vierteljahrsschrift für wissenschaftliche Philosophie*, 16 (2), 1892a.
- FREGE, G. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100 (1), pp. 25-50, 1892b.
- FREGE, G. Logik. In: H. HERMEL; KAMBARTEL, F.; KAULBACH, F. (Hrsg.) (1969), pp. 137-63, 1897.
- FREGE, G. Einleitung in die Logik. In: H. HERMEL; KAMBARTEL, F.; KAULBACH, F. (Hrsg.) (1969), pp. 201-12, 1906a.
- FREGE, G. Kurze Übersicht meiner logischen Lehren. In: H. HERMEL; KAMBARTEL, F.; KAULBACH, F. (Hrsg.) (1969), pp. 213-18, 1906b.
- FREGE, G. Brief an Husserl. In: GABRIEL, G. et al. (Hrsg.) (1976), pp. 101-6, 1906c.
- FREGE, G. Der Gedanke. Eine logische Untersuchung. *Beiträge zur Philosophie des deutschen Idealismus*, 2, pp. 58-77, 1918-19.
- FREGE, G. *Nachgelassene Schriften*. H. HERMEL; KAMBARTEL, F.; KAULBACH, F. (Hrsg.). Hamburg: Felix Meiner, 1969.
- FREGE, G. *Posthumous Writings*. HERMES, H.; F. KAMBARTEL; F. KAULBACH (eds.). Oxford: Blackwell, 1979.
- FREGE, G. *Wissenschaftlicher Briefwechsel*. GABRIEL, G. et al. (Hrsg.). Hamburg: Felix Meiner, 1976.
- GUTZMANN, D. Expressives and Beyond: An Introduction to Varieties of Use-Conditional Meaning. In: GUTZMANN; H.-M. GÄRTNER (eds.) (2013), pp. 1-58.
- GUTZMANN, D. *The Grammar of Expressivity*. Oxford: Oxford University Press, 2019.
- GUTZMANN, D.; GÄRTNER, H.-HM (Eds.). *Beyond Expressives: Explorations in Use-Conditional Meaning*. Leiden: Brill, 2013.
- HAY, R. Hybrid Expressivism and the Analogy Between Pejoratives and Moral Language. *European Journal of Philosophy*, 21 (3), p. 450-474, 2013.
- HOM, C. Pejoratives. *Philosophy Compass*, 5 (2), pp. 164-185, 2010.
- HORN, L. R. Towards a Fregean Pragmatics: Voraussetzung, Nebengedanke, Andeutung. In: I. KECSKES; L. R. HORN (Eds.) (2007), pp. 39-69.
- HORN, L. R. I Love Me Some Datives: Expressive Meaning, Free Datives, and F-Implicature. In: D. GUTZMANN; H.-M. GÄRTNER (eds.) (2013), pp. 143-189.
- KAPLAN, D. The Meaning of 'Ouch' and 'Oops': Explorations in the Theory of Meaning as Use. *Cornell Conference on Context Dependency*, pp. 1-29, 1999.

KAPLAN, D. The Meaning of 'Ouch' and 'Oops'. Howison Lecture in Philosophy delivered at UC Berkeley. Transcribed by E. Coppock, 2004.

KARTTUNEN, L. Presuppositions of Compound Sentences. *Linguistic Inquiry*, 4 (2), pp. 169-193, 1973.

KECSKES, I; HORN, L. R. (Eds.) *Explorations in Pragmatics: Linguistic, Cognitive, and Intercultural Aspects*. Berlin, New York: De Gruyter, 2007.

KORTUM, R. *Varieties of Tone: Frege, Dummett, and the Shades of Meaning*. London: Palgrave Macmillan, 2013.

KURASUGI, K.; STANTON, R. (eds.). *Philosophy and Linguistics*. London, New York: Routledge, 1999.

NEALE, S. Coloring and Composition. In: Kurasugi, K.; R. Stainton (eds.) (1999), pp. 35-82.

NEALE, S. Implicature and coloring. In: G. Cosenza (ed.) (2001), pp. 135-180.

ORLANDO, E.; SAAB, A. (eds.). *Slurs and Expressivity: Semantics and Beyond*. Lanham, MA: Lexington Books, 2020.

PICARDI, E. Colouring, Multiple Propositions, and Assertoric Content. *Grazer Philosophische Studien*, 72 (1), pp. 49-71, 2006.

PICARDI, E. On Sense, Tone, and Accompanying Thoughts. In: R. E. AUXIER; HAHN, L. E. (eds.) (2007), pp. 491-520.

POTTS, C. The Expressive Dimension. *Theoretical Linguistics*, 33 (2), pp. 165-198, 2007.

SANDER, T. Two Misconstruals of Frege's Theory of Colouring. *The Philosophical Quarterly*, 69 (275), pp. 374-92, 2019.

SANDER, T. Fregean Side-Thoughts. *Australian Journal of Philosophy*, 99(3), pp. 455-71, 2021.

SANDER, T. Taxonomizing Non-At-Issue Contents. *Grazer Philosophische Studien*, forthcoming.

SAUERLAND, U. Beyond Unpluggability. *Theoretical Linguistics*, 33 (2), pp. 231-6, 2007.

SOSA, D. (ed.). *Bad Words: Philosophical Perspectives on Slurs*. Oxford: Oxford University Press, 2018.

SOUTIF, L.; MÁRQUEZ, C. Expressives and the Theory of Bias. In: ORLANDO, E.; A. SAAB (eds.) (2020), pp. 95-116.

VALTONEN P. Semantic Innocence and Kaplanian Inferences. *Sats*, 20 (1), pp. 19-33, 2019.