ORIGINAL ARTICLE

ANALYTIC PHILOSOPHY WILEY

Towards a Fregean psycholinguistics

Thorsten Sander

Fakultät für Geisteswissenschaften,
Institut für Philosophie, Universität
Duisburg-Essen, Essen, Germany

Correspondence
Thorsten Sander, Fakultät für
Geisteswissenschaften, Institut für
Philosophie, Universität Duisburg-Essen,
D-45117 Essen, Germany.
Email: thorsten.sander@uni-due.de

Abstract

This paper is partly exegetical, partly systematic. I argue that Frege's account of what he called "colouring" contains some important insights on how communication is related to mental states such as mental images or emotions. I also show that the Fregean perspective is supported by current research in psycholinguistics and that a full understanding of some linguistic phenomena that scholars have accounted for in terms of either semantics or pragmatics need involve psycholinguistic elements.

1 | INTRODUCTION

Frege's impact on contemporary philosophy of language is second to none, but there is one aspect of his account of human communication that has either been ignored or almost universally rejected as flawed. When Frege discusses items that express "side-thoughts" or add "colouring" to a sentence,[1] he frequently (though not universally) characterizes the contribution of such items in terms of the mental states that are evoked by them. The word "cur", for instance, is on Frege's view tied to the mental image of a scruffy dog (Frege, 1897: 152). Similarly, the sentence "Napoleon, who recognised the danger to his right flank, himself led his guards against the enemy position" is for Frege psychologically connected to the side-thought that Napoleon's recognition of the danger was the reason for leading his guards against the enemy position (Frege, 1892: 47). More generally, he claims that many linguistic items are "meant to act on the feelings and mood of the hearer, or to arouse his imagination" (Frege, 1918: 63, 1997: 330).

Given Frege's unabashed anti-psychologism, such claims may at first glance seem surprising, but of course there is no tension (let alone a contradiction) involved here. His anti-psychologism concerns the realm of logic and, more specifically, the logical contents which he called "thoughts" (*Gedanken*). Frege's idea that thoughts ain't in the head is perfectly compatible with the assumption

---

[1]Compare Dummett (1981: 83–89), Picardi (2007), Horn (2007, 2013), Williamson (2009: 148–154), Sander (2019, 2021).

that more subtle shades of meaning are accountable for in terms of human psychology. His idea that psychology plays an essential role in a full account of human communication gives him a unique place in the history of the philosophy of language: up to very recently, almost all prominent scholars in that field have assumed that we simply ought to ignore things such as mental images.

Frege's logical (and thus restricted) anti-psychologism turned, in the hands of later authors, into a more or less unrestricted and rather dogmatic anti-psychologism, according to which a theory of human communication can safely ignore what is going on in our minds. This unrestricted anti-psychologism is clearly visible in some recent discussions of Frege's notion of colouring. For instance, Williamson (2009: 149) has claimed that "Frege discusses his examples [such as "cur"] by speaking unhelpfully of the images and feelings that the words evoke in hearers." Why "unhelpfully"? On Williamson's view, what needs explaining is that speakers "are not willing to assert 'Lessing was a Boche,' even on reflection, unless they are xenophobes", and his "natural" answer to the question of why this is so has it that the sentence "Lessing was a Boche" conventionally implicates that Germans are cruel (Williamson, 2009: 149). Williamson seems to assume, then, that the Gricean notion of conventional implicature can account for the stable use of pejoratives, while Frege's talk about mental images or feelings cannot. We shall see below that this assumption is far from being self-evident.

Here is a second example. When Picardi discusses Frege's idea that certain sentences give rise to what he calls "side-thoughts" (more on which in Section 4), she notes:

> In commenting on the material conditional Frege employs the notion of [side-thought] to indicate a psychological association [...] which invariably accompanies the production and understanding of the "if-then" construction. [...] That the triggering of such *Nebengedanken* [...] may be governed by psychological laws was a concession to the scientific psychology of his day.
>
> (Picardi, 2007: 502)

Picardi does not really say that Frege's account is flawed, but by talking about *concessions* and the psychology *of his day*, she strongly suggests that Frege gave too much credit to an obsolete and discredited kind of psychological theory. But is it really obsolete? In his book *Thinking, fast and slow*, Kahneman has described our mind (more specifically, the part of our mind that is responsible for "fast" thinking) as an "associative machine", and he offers a nice example of the associations words may evoke: reading, and understanding, the two words "Bananas" and "Vomit", simply printed next to each other, will have all sorts of effects on us, including a "temporary aversion to bananas" (2011: 50).

So the idea that understanding the workings of our mind has to take associations into account does not seem to be part of an outdated psychological theory. This idea, moreover, accords with introspective evidence, available independently of elaborated experiments concerning priming effects: that words frequently do something with us, by evoking certain concepts, mental images or emotions is an experience familiar to any language user.

At this point of the dialectic, it may seem tempting to downplay the role of such associations by conceding that they exist and, at the same time, claiming that they are irrelevant to understanding human communication: Yes, Frege (1897: 151, 1997: 40) was right that "[a]nyone who hears the word 'horse' and understands it will probably have straightaway a picture of a horse in his mind", but so what? That mental image has, as Frege stresses, nothing to do with the sense of the word "horse", and since it is subjective, it clearly cannot even play a marginal role in understanding human communication.

Such a line of thought lies behind Dummett's influential attack on Frege's account of what Dummett has called "tone", i.e. the kind of meaning conveyed by items such as "cur":

> [Frege] accounts for tone as a matter of the association with a word or expression of certain 'ideas' (*Vorstellungen*), by which he means mental images. This is not a particularly plausible explanation of the phenomenon [...]. Frege makes a poor explanation worse by suggesting that mental images are incommunicable in principle: no two people can ever know that they have the same mental image. It would follow that tone was a feature of meaning which was, in principle, subjective. This conclusion is a simple contradiction. Meaning, under any theory whatsoever, cannot be *in principle* subjective, because meaning is a matter of what is *conveyed* by language.
>
> (Dummett, 1981: 85).

That meaning can concern only what is conveyed by a linguistic item seems to be a near-consensus in current theorizing. Most scholars would assume that meaning (broadly construed) comes in exactly two flavours, speaker meaning and linguistic meaning, and both flavours have to do with what is conveyed by an utterance. In the latter case, a content is conveyed because it is encoded; and in the former case, it is conveyed because it is non-naturally meant by the speaker. But is it really self-evident that "meaning" cannot be analysed in an alternative way? Consider how Stevenson (1944: 54) defines an essentially psychological notion of meaning:

> The meaning of a sign, in the psychological sense required, is not some specific psychological process that attends the sign at any one time. It is rather a dispositional property of the sign, where the response [...] consists of psychological processes in a hearer, and where the stimulus is his hearing the sign.

What makes Stevenson's idea particularly interesting here is that his notion of meaning may be construed as an audience-directed counterpart to a specifically Gricean notion of linguistic meaning.[2] Griceans will typically define the notion of linguistic meaning in terms of what speakers *regularly* speaker-mean by an item (cf. Grice, 1989: 350), and Stevenson similarly defines "the meaning of a sign" in terms of its *regular* effects on an audience. Such an audience-directed notion of meaning, mirroring the speaker-directed notion, may be helpful to describe cases where the impact of a word is independent of either its encoded meaning or of what a speaker intended to convey by it. The phenomenon of "unintentional dogwhistles" (Saul, 2018: 367–371) is, arguably, a salient example. So contrary to what Dummett suggests, there is an intelligible notion of meaning that goes beyond what is "conveyed" by certain items, and we shall also see that Dummett's above argument involves a non-sequitur: a public communicative phenomenon may be explainable in terms of something that is essentially private.

The incredulous stare Frege's partly subjectivist account of "tone" has received from philosophers such as Dummett, Picardi, and Williamson might be explained by the fact that Frege's psycholinguistic perspective is irrelevant, and indeed alien, to most of the questions that were

---

[2]Compare also Saul's (2002: 242) notion of "audience-implicature", which can be generalized to a broader notion of "audience-meaning".

once construed as the big questions in the philosophy of language. If you are mainly concerned with questions such as how language "hooks onto the world", then you may safely ignore Frege's ideas.[3] However, philosophy of language has significantly changed over the last two decades. Philosophers, as well as linguists, have attended to various sorts of non-at-issue contents on the one hand, and to all sorts of potentially pernicious language on the other, and accounting for some of these phenomena may essentially involve the *psychological effects* ("ψ-effect", for short) that words may have on us.[4]

So Frege's account of tone may be not as misguided as many scholars have thought, and in this paper I will argue that it contains a number of important insights. In what follows I will use some of his ideas as a starting point and relate them to current empirical and theoretical research. The leitmotif that will emerge in this paper is that certain phenomena that have been accounted for by semantic or pragmatic mechanisms such as presupposition or implicature ought to be explained (partly) in psycholinguistic terms.

Here is how I shall proceed: I begin, in Section 2, with some remarks on Stanley's account of "social meaning" and on Cappelen's notion of lexical effect. In Section 3, I then turn to what Frege says about the contribution of "cur". On my view, his account offers an important insight on how the expressiveness of certain items may interact with the mental images evoked by them. Section 4 deals with Frege's account of accompanying side-thoughts. I argue that a Fregean account can explain coherence relations between sentences more successfully than an account that relies on the notion of conversational implicature, and I also show how an intersubjective communicative phenomenon may be explainable by something that is subjective. In Section 5, I turn to the non-arbitrary shades of meaning that are evoked by the mere sound-quality of certain words, and in Section 6, I briefly discuss a (non-Fregean) example where a ψ-effect may be a matter of grammar. In the final section, I briefly propose a taxonomy of ψ-effects.

2 | LEXICAL EFFECTS AND SOCIAL MEANING

While philosophers such as Dummett assume that (to put it mildly) psycholinguistic considerations ought not to play a large role in the philosophy of language, some other scholars have more recently (and, in part, reluctantly) argued to the contrary.[5]

In his book on propaganda, Stanley has argued that the contribution of "code words" (or "dog whistles") such as "welfare" or "inner city" may be partly explained by the fact that propagandists associate words with images. One of his examples comes from Klemperer's (2013: 1–8) account of how national-socialist propaganda shifted the meaning of the word "heroism":

> The media associated [the word "heroism"] with specific images: the racecar driver, the tank commander, the Storm Trooper. The images the media associated with [the word] became part of the social meaning of the term "heroism" for those raised

---

[3]A nice example of such a focus is Devitt's and Sterelny's (1999) book. The subtitle promises an introduction to *the* philosophy of language, but as the main-title indicates, it is mainly concerned with how language mirrors reality. Accordingly, most issues treated in current pragmatics are largely ignored.

[4]My notion of psychological effect is modelled on Cappelen's (2018: ch. 11) notion of lexical effect, more on which in the next section.

[5]Several relevance theorists have argued for the relevance of "non-propositional effects"; cf. Carston (2018) and Wilson and Carston (2019). See also Stojnić and Lepore (2022) who have argued that slurs may be accounted for by the associations triggered by (articulations of) them. Cf. Lepore and Stone (2018).

under National Socialism. As with Frege's description of the images associated with the term "cur," it was impossible for those raised under National Socialism not to have the word evoke those images. The Republican "Southern Strategy" was to associate the language surrounding social welfare programs with images reinforcing the stereotype of urban Blacks as lazy.

<div align="right">(Stanley, 2015: 158)</div>

Up to this point, the explanation is purely psycholinguistic: a word may be used repeatedly along with *actual* images in the mass media, thereby acquiring an enriched social meaning such that the word has the power to evoke *mental* images. This would mean that the contribution of code words cannot be captured by any mechanism standardly discussed in either semantics or pragmatics.

Stanley, however, misses the full potential of this insight in that he limits the role of ψ-effects to explaining how a word such as "welfare" gets associated with a particular non-at-issue content. Once that meaning change has taken place, ψ-effects no longer play a role; when using the word "welfare", speakers simply convey an additional propositional content:

> When the news media connects images of urban Blacks repeatedly with mentions of the term "welfare," the term "welfare" comes to have the not-at-issue content that Blacks are lazy. At some point, the repeated associations are part of the meaning, the not-at-issue content.

<div align="right">(Stanley, 2015: 158)</div>

The notion of non-at-issue content[6] to which Stanley appeals here goes back to Potts's (2005) account. One of Potts's standard examples of non-at-issue content (or more specifically of what he calls "conventional implicature"), are expressives such as "that bastard". A sentence such as "That bastard Conner got promoted" (Potts, 2005: 157) has the at-issue content that Conner got promoted, coupled here with a non-at-issue content concerning the speaker's attitude towards Conner. Now one crucial property of non-at-issue contents is that they cannot be challenged in the same way as at-issue contents. Compare two possible reactions to a sentence involving the word "bastard":

(2-1) A: That bastard Conner got promoted.

(2-1a) B: That's not true, Conner is not a bastard.

(2-1b) B: Wait a minute, Conner is not a bastard.

It is widely acknowledged that by employing items such as "wait a minute" one is able to target various kinds of secondary contents that cannot be directly denied, which is why (2-1b) is a felicitous reply to (2–1), while (2-1a) is not.[7] Now if the term "welfare" indeed conveyed a non-at-issue content proper, then we should expect the following discourse to be natural:

---

[6]Many scholars, including Stanley, talk about "not-at-issue contents". For aesthetic reasons, I follow Horn (2013: 64), and some others, in employing the term "non-at-issue" here.

[7]I follow Potts (2015: 174–5) in assuming here that "wait a minute"-style devices "can be used to object to a wide range of non-at-issue content, including conventional implicatures, appropriateness conditions, and conversational implicatures." Compare also Salmon (2011: 3426) and Tonhauser et al. (2013: 81), who have advanced essentially the same claim.

(2–2) A: Social welfare spending has decreased relative to GDP.

(2-2a) B: Wait a minute, Blacks are not lazy.

But it is not. (See Saul [2018: 374] for a similar observation.) This is not to deny that in many contexts, but plausibly not in the context of (2–2), the term "welfare" sneakily communicates racist beliefs or attitudes, but it seems to do so in an even more "backgrounded" manner than is the case with ordinary non-at-issue contents.[8] And the assumption that code words trigger associations, but do not encode contents offers a possible explanation for that.

Stanley only reluctantly acknowledges the relevance of psycholinguistics: associations offer an explanation for why words such as "welfare" *come* to encode certain contents, but once that has happened, mental images and the like seem communicatively inert on his account. Cappelen, in contrast, has argued that the study of what he calls "lexical effects" is "potentially enormously significant" (2018: 130). By that term he means all sorts of cognitive or non-cognitive effects words may have on us. On his view, names (especially brand names), pejoratives, and metaphors are phenomena whose explanation must essentially involve such effects, and Cappelen (together with Dever) also argues, in contrast to Stanley, that such effects may directly account for the specific punch of code words:

> [The non-cognitive lexical effects carried by items such as "inner city"] are not contents: they trigger pictures, memories, affect your mood, your motivation, and can change 'the way you think about something,' rather than the content of what you think.
>
> (Cappelen & Dever, 2019: 119)

I largely agree with Cappelen (and Dever), but in one respect he still seems to underestimate the significance of lexical effects. He makes two conjectures about such effects which, if taken at face value, would provide a rationale for largely ignoring them. On Cappelen's view, lexical effects are *not* stable "across large populations", and they also "change easily over time" (Cappelen, 2018: 130).

If the term "lexical effect" covers *any* psychological effect words may have, then it is true of course that a large portion of them will be highly idiosyncratic. For instance, the fact that Rigby Reardon (from the movie *Dead Men Don't Wear Plaid*) shows a strong emotional reaction to the word "cleaning woman", but not to its German counterpart "Reinemachefrau", may be useful for predicting Rigby's behaviour, but it is a fact we may safely ignore as philosophers of language.

In other cases, psychological effects may be shared among all members of a particular culture, or even among all human beings,[9] and such effects may be essential to understanding the overall significance of certain items. (We shall encounter several examples in the next sections.) But even if there are some differences in the effects, this may not matter much to the communicative impact of a particular item. Consider Frege's (1897: 152) claim that the

---

[8]Note that (2-2a)-style of replies seem odd even when it is more or less obvious that a speaker intentionally used "welfare" as a code word. In more recent work (jointly authored with David Beaver), Stanley has offered a more nuanced account of dogwhistles and other kinds of non-at-issue contents (Beaver & Stanley, 2019). But I think it is instructive to see where exactly Stanley's 2015 account goes awry.

[9]Kahneman's example from above ("bananas"/"vomit") seems to be (almost) universal.

contribution of the word "cur" may be partially accounted for by mental images of a scruffy dog (more on which in the next section). That word may evoke in me a mental image closely resembling a recent ugly dog meme, while you may think of the surly dog depicted in Sir Edwin Landseer's 1829 painting *Low Life*. But such differences may be ultimately irrelevant since both mental images belong to the same type and since, moreover, both images may in turn lead to essentially the same aesthetic and emotional reactions. They are, in other words, functionally equivalent.

Consider another example. In 2018, a German clothing store published an advertisement showing a couple of white men's shirts, supplemented with the German slogan "Jedem das Seine"—a literal translation of the Latin phrase "suum cuique", "to each his own".[10] The communicative intention in using that slogan seems obvious, but the ad caused a controversy because of a particular ψ-effect. Since the slogan was shown in the main gate of the Buchenwald concentration camp, most Germans will take it to be tainted by National Socialism. More specifically, it may be associated with a particular concentration camp and may often evoke a blurry mental image of the gate. This is an example of an explanatory and culturally stable ψ-effect, even though the specific effects may vary among different subjects. Explaining the fact that the German phrase is usually avoided in Germany just requires that it is reliably associated with National Socialism. It does not matter whether that association specifically involves mental images or cognitive states.[11]

3 | FREGE ON EXPRESSIVES AND MENTAL IMAGES

Let us turn to what Frege says about the expressive contribution of the word "cur". While Frege's example has become so popular that it is frequently not even attributed to him, his nuanced analysis of it has often been misrepresented as a simple kind of Ayer-style expressivism. According to Williamson's (2009: 149) reading of Frege, for instance, "cur" simply "conveys an attitude of contempt for the dog on the part of the speaker". And while it is true of course that this is an important part of the Fregean account, "cur" involves, on Frege's view, more than just the expression of an attitude.

Now unfortunately, the translation of Frege's posthumous writings, published more than 40 years ago, is sometimes not particularly faithful to the original. So here is a new translation of the pertinent passage of Frege's 1897 logic:

> [1] If we compare the sentences "This dog howled the whole night" and "This cur howled the whole night", we find that thought is the same. [2] We learn neither more nor less from the first sentence than from the second. [3] But while the word "dog" is related indifferently to pleasure and displeasure (*Lust und Unlust*), the word "cur" is related much more closely to displeasure and thereby (*damit*) gives a hint (*Wink*) to imagine (*vorstellen*) the dog as being somewhat scruffy (*ruppig*). [4] Now even if this does a great injustice to him, one cannot say that the

---

[10]Cf. https://en.wikipedia.org/wiki/Jedem_das_Seine.

[11]In this respect, ψ-effects are perhaps not very different from the contribution of ordinary words. Compare Ludlow (2014), who argues (successfully, I think) that our lexicon is much less "static" than scholars usually assume. If Ludlow is right, then the content actually shared by interlocutors may at best concern a highly indeterminate core meaning.

second sentence is therefore false. [5] Anyone who utters it indeed expresses (*äussert*) a certain disdain (*Geringschätzung*); but this does not belong to the thought expressed.

(Frege, 1897: 152)

The main differences between my translation and the one in Frege's *Posthumous Writings* (reprinted in Beaney's *The Frege Reader*) concern [3] and [5].[12] In the translation by Long and White, the first part of [5] is rendered as "anyone who utters this sentence speaks pejoratively" (1979: 140). Frege, however, is not concerned here with *how* persons speak, but instead with *what* they express when they are speaking. The standard translation of [3] is similarly misleading: Frege simply does not talk about (un)pleasant *associations* here. By using the two terms "pleasure" and "displeasure", Frege rather alludes to 18th and 19th century discussions of the two feelings that were construed as being intrinsically linked to the conative faculty, claiming that "cur" is related to displeasure.[13] He does not specify the kind of relation any further here, but in [5] he clearly suggests that (a specific kind of) displeasure is what is expressed by "cur". And in contrast to what is suggested by the standard translation (1979: 140), the word "cur" does not merely "imply" that the speaker feels a kind of disdain. Frege rather says that the negative attitude seems to lie in the word ("die Geringschätzung [...], die in dem Worte 'Köter' zu liegen scheint"; 1897: 152), which may suggest that he construed its expressive function as part of its linguistic meaning.

The picture that has emerged so far seems indeed broadly expressivist: by using the word "cur", speakers express their negative attitudes (perhaps as a matter of linguistic meaning), but in [3], another element comes into play: "cur" is also related to the mental image of a scruffy dog. According to the standard translation, the word "cur" "puts us rather in mind of a dog with a somewhat unkempt appearance" (Frege, 1979: 140), where the "us" suggest that "cur" behaves symmetrically with respect to speaker and hearer. But this is not Frege's idea. Frege rather seems to say that "cur" is used with the intention of evoking a mental image *in the addressee*, while the speaker expresses a conative attitude when using that word. Note that Frege employs here the term "hint" (*Wink*), which he already used before in the same paragraph when discussing the relation between "walk" and "stroll": that languages contain such words that make no difference to the thought expressed is "useful" precisely because they enable a speaker to send hints such that these words "can act in different ways on the feelings and imagination *of the hearer*" (Frege, 1897: 151f, 1979: 140, emphasis added). As is suggested by the term "hint", hints concern what is happening in the mind of an addressee.

On Frege's view, the mental image of a scruffy dog and the disdain expressed by the speaker are not independent. In [3], he seems to claim that the hint as to the mental image is somehow grounded in the disdain. This is surprising. He does not tell us anything specific about the nature of that grounding relation, but I think he may have been up to something important here. And his insight may have to do with how we can affect the attitudes of our interlocutors.

---

[12]Compare Frege (1979: 140): "[3] But whilst the word 'dog' is neutral as between having pleasant or unpleasant associations, the word 'cur' certainly has unpleasant rather than pleasant associations and puts us rather in mind of a dog with a somewhat unkempt appearance. [...] [5] True, anyone who utters this sentence speaks pejoratively, but this is not part of the thought expressed."

[13]When Kant (1991: 40; AA VI 211) discusses the "capacity for desire" (*Begehrungsvermögen*), he notes that "pleasure (*Lust*) or displeasure (*Unlust*) [...] is always connected with desire or aversion".

It is sometimes claimed that non-at-issue contents may be particularly pernicious because they have the force to "smuggle" highly controversial material into the common ground. According to Stanley (2015: 135), the "not-at-issue content of an utterance is not advanced as a proposal of a content to be added to the common ground. Not-at-issue content is *directly* added to the common ground." And Potts has advanced essentially the same claim specifically with respect to expressives such as "that bastard": when Anne utters "That bastard Conner got promoted", then on Potts's (2005: 157) view, "the contribution of the epithet *that bastard* as used by Anne does become part of the common ground."[14] Now the first thing to ask here is whether the notion of common ground can be applied to expressive contents in the first place. If, as Stalnaker (2002: 706) put it, "the common ground of a conversation is just what is common belief among the participants in a conversation", then it is not clear what exactly is added to it when a speaker uses an epithet. On the assumption that Anne is speaking truly *and felicitously*, her utterance of.

> (3–1) That bastard Conner got promoted

will indeed have the effect that her interlocutors come to believe that Anne does not like Conner (or the like). But the propositional content *that Anne does not like Conner* seems to be different from the emotive content that is conveyed by (3–1). (In metaethical terms: expressivism is not the same as speaker subjectivism.) One might of course devise a broader notion of common ground that covers not only the doxastic, but also the conative states shared by the interlocutors. (Let us call this "common-ground-plus" or "CGP"). It does not seem plausible, however, that expressive non-at-issue contents generally have a direct effect on the CGP. If you think that Conner is a really nice guy, then why should my uttering (3–1) change your attitude towards Conner?[15] (My utterance, however, may have the effect that you change your attitude towards me.)

At this point, Frege's implicit insight comes into play. Though it may be impossible, or difficult, to change your interlocutors' attitudes simply by expressing them, certain words may have reliable *indirect* effects on your addressees. If "cur" is not only a means of expressing your disdain, but also of evoking particular ideas in the mind of a hearer, then these ideas may in turn affect the hearer's conative attitudes. Having a mental image of an aesthetically revolting dog may lead to a change in one's conative attitude towards it. So we may read Frege as claiming that your disdain for a particular dog will give you a reason for choosing a term which (via a mental image evoked by it) may make your addressee share your attitude.

This insight of Frege's is foreshadowed in ancient rhetoric. There is a consensus among rhetoricians that actually persuading an audience will often require eliciting an emotional response, which raises the question of how that might be achieved. A rather straightforward way of evoking emotions is recommended by Quintilian, who notes how our "vice" of forming mental images (for example, in daydreams) can be exploited: a detailed description of an event such as a

---

[14]Note that both Stanley and Potts make a claim that goes beyond the fairly uncontroversial observation that non-at-issue contents become part of the common ground, unless they are challenged (cf. Stalnaker, 2014: 47). Stanley (2015: 157) has it that "even the act of raising the expressions to salience" by objecting to them conveys "the negative social meanings". And on Potts's (2005: 158) view, expressives are essentially unchallengeable in that "outright denials of their content by a hearer will make little sense".

[15]However, there is the phenomenon of emotional "contagion", i.e., "the tendency to automatically mimic and synchronize facial expressions, vocalizations, postures, and movements with those of another person and, consequently, to converge emotionally" (Hatfield et al., 1992: 5).

murder may lead, via mental images, to essentially the same emotive response that seeing the event itself would have caused. Quintilian aptly called this an example of "showing" something:

> The result [of such a description] is ἐνάργεια (vividness), [...] which does not seem to be talking (*dicere*) but rather showing (*ostendere*). And the affects will follow as though we were experiencing (*intersimus*) the things themselves.
>
> (Quintilian 1854: 263; *Institutio oratoria*, VI 2, 32, my translation)

That longer stretches of discourse may have an emotional impact on us seems evident (think, for example, of reading a horror story),[16] but is it plausible that, as Frege seems to assume, the same mechanism is operative in the case of single words such as "cur"? The answer is yes, at least in principle: there is ample psycholinguistic evidence that some words are "emotion-laden". (See Foolen [2015] for a helpful survey.) Words like "coffin", "rape" or "debt" are emotive insofar as they refer to things that are emotionally significant to us.

Now of course Frege may have been wrong about the exact contribution of the German word "Köter", but that does not mean that he was altogether wrong about the communicative effects of mental images. Recall Cappelen's and Dever's (2019: 119) above-quoted claim that code words trigger things such as images. Obviously, the exact working of such words will involve complex theoretical and empirical issues, and so we should tread carefully here. But there are some observations that may support the claim that the use of code words essentially involves the exploitation of lexical effects. Consider an experiment by White (2007), nicely summed up by Khoo (2017: 36):

> [W]hen food stamp programs are described as benefiting "inner-city" families, white racial prejudice toward African Americans is significantly more predictive of increased opposition to government spending for food stamp programs than when the same food stamp programs were described as benefiting "African American families," "poor families," or just "American families".

It is generally acknowledged that in the Unites States "inner city" functions as a code word for African Americans, which explains why it makes a difference whether one talks about "violent criminals" or about "violent *inner city* criminals" (cf. Saul 2018: 367). Now Khoo (2017) has offered a "simple theory" of code words that is not semantic but instead inferential. His account involves an utterance, a preexisting belief, and an inference:

(a) Politician Z utters: "The food stamp program will primarily benefit inner-city Americans".
(b) Z's audience believes that the inner city is mostly populated by poor African Americans.
(c) Z's audience infers that the food stamp program will primarily benefit poor African Americans.[17]

Khoo's theory can indeed explain the difference between "inner-city families" and "American families", but according to the above experiment there is also a difference between "inner-city families" and "African American families". If an apparently non-racial statement, as in (a), triggers racial prejudice via an inferential process, then why is such an effect absent in the case of an *explicitly* racial statement such as "The food stamp program will primarily benefit African Americans"?

---

[16]Compare the kind of propaganda that involves spreading rumours about a particular group. Surprisingly, this wide-spread kind of "bad language" has received virtually no attention from philosophers.

[17]The example is due to Khoo (2017: 47). I've slightly modified the presentation.

As an alternative (or as a supplement), one might claim that the difference is (partially) explainable in terms of ψ-effects. As said, this is an issue that cannot be settled by mere armchair reflection, but could it not be that the word "inner city" does not only trigger stereotypical beliefs (as Khoo assumes), but additionally has all sorts of non-cognitive effects on an addressee?[18] That word may evoke the mental image of an unkempt place which in turn may lead to an aesthetic response; it may, moreover, be associated with concepts such as *crime* or *drugs*; and it may (therefore) also be an "emotion-laden word".

4 | SIDE-THOUGHTS, MEANING$_{NN}$ AND MANIPULATION

In Section 2, I claimed that, pace Dummett, an intersubjective communicative phenomenon may be explainable by something that is essentially subjective. This needs to be explained. So let us start with an analogy. In metaethics, it is commonly assumed that "thick terms" mix evaluative and descriptive content. On such a view, the negative evaluation that appears to be carried by words such as "lewd" or "brave" would be part of their linguistic meaning, and accordingly thick terms would be either triggers of lexical presuppositions or items carrying conventional implicatures. There is, however, an alternative view. One might argue that the contribution of thick terms is exclusively descriptive: what appears to be an evaluative aspect of meaning is, on such a view, just an evaluative judgment reliably triggered by purely descriptive contents. For instance, "lewd" might simply *mean* something along the lines of "promiscuous or sexually explicit", and the negative evaluation "communicated" by that word may be explained by the attitudes people have as a matter of fact.

Such an account of thick terms has been defended by Blackburn (1992) and, more recently, by Väyrynen (2013). Blackburn offers a nice analogy:

> We might expect someone who talks of a house as containing south facing windows to be implying or inviting a favourable attitude to that feature, yet 'contains south facing windows' is not usually thought of as a thick term, and certainly there is no linguistic convention that a house with south facing windows should be favourably regarded. Apart from anything else, there would be no need for such a convention, given what people normally desire.
>
> (Blackburn, 1992: 287)

In the last sentence, Blackburn parenthetically makes an important point: a subjective mental state such as a desire may have much the same effect as a linguistic convention, provided that it is a state people "normally" have. As far as communicative effects are concerned, an estate agent may, as it were, perform a speech act of recommending a house simply by asserting that the house has a certain property *F*—if potential buyers normally want a house that is *F*.

Note that this kind of communication works even on the Fregean assumption that mental states are, in some sense, essentially private. In his "Über Sinn und Bedeutung", Frege notoriously claims that an "exact comparison" between ideas is impossible "because we cannot have both ideas together in the same consciousness" (1892: 30, 1997: 155). But an *exact* comparison is not what is at issue here. If, for instance, people are prepared to pay more for a house that is *F* than for an otherwise comparable house that is not *F* (or if they simply tell you that they like

---

[18]When applying his account to examples of what Haslanger (2013) has called "social meaning", Khoo (2017: 57) concedes that "a full account" of this phenomenon should include various kinds of non-cognitive states.

houses that are *F*), then this all you need to know about their desires here. You do not have to know what exactly these people feel from the inside when they are desiring something.

I will remain non-committal about thick terms here, but Blackburn's example makes plausible the idea that the social significance of an item may be (partly) explainable in terms of individual mental states. Let us now turn to a Fregean example where ψ-effects may be important too. Frege claims that an utterance of.

> (4–1) Napoleon, who recognized the danger to his right flank, himself led his guards against the enemy position.

> (1892: 47, 1997: 168)

may be used to convey that Napoleon's recognition of the danger was the reason why he led his guards against the enemy position. On Frege's view, (4–1) is an example of what he called an accompanying side-thought (cf. Horn, 2007; Picardi, 2007; Sander, 2021). In such cases, a thought that is not expressed by a sentence is associated with it "according to psychological laws" (Frege, 1892: 46, 1997: 168). And it seems as though Frege was right here: when we are being told that (i) a person *P* performed an action *A* and that (ii) *P* was in a mental state *M*, then we often (though not universally) will hear (ii) as a possible explanation for (i). Note that this is independent of the syntax of (4–1). Two adjacent sentences ("Napoleon recognized the danger to his right flank. He led his guards against the enemy position") will often be read in exactly the same way.

Such psychological enrichment may be operative even in cases where what is said appears textually incoherent. Consider an example:

> (4–2) John took a train from Paris to Istanbul. He likes spinach.

> (Hobbs, 1979: 67)

Hobbs presents (4–2) as a counterexample to the claim that a "discourse is coherent *because* successive utterances are 'about' the same entities" (Hobbs, 1979: 67), and he is clearly right that coherence involves more than just co-reference between sentences. But he himself notes that, contrary to first appearance, (4–2) may turn out to be coherent if one adds an explanation along the following lines: "Well, maybe the French spinach crop failed and Turkey is the only country ..." (ibid.) That we are willing to engage in such speculations about John's motives confirms Pinker's (2014: 141) observation that "whenever one sentence comes after another, readers need to see a connection between them. So eager are readers to seek coherence that they will often supply it when none exists."

It seems undisputable that there is such a desire for coherence, but this fact in itself does not yet tell us anything about the mechanism operative here. Frege claimed that, in some cases at least, the mechanism is that of psychological association, but there is another option. Alternatively, the enrichment that takes place in cases such as (4–1) or (4–2) might be due to conversational implicature, which would show that the standard toolbox of pragmatics offers the tools necessary for treating such cases. Now Grice's original account of conversational implicature (henceforth I shall drop the qualification "conversational") operates at the level of whole utterances, and so it is unclear whether it can be extended to subsentential clauses, as in (4–1), and to sequences of sentences, as in (4–2). In contrast, Levinson's (2000) neo-Gricean theory of generalized conversational implicatures purports to be applicable at least to examples such as (4–2). (As far as I can see, Levinson never mentions examples like (4–1).) On Levinson's view, some examples of parataxis such as

(4–3) John fell and broke his leg. He lost his grip on the cliff.

(Levinson, 2000: 124)

give rise to a causal enrichment: the second sentence of (4–3) will be read as offering an explanation for the event described in the first sentence.

Levinson's account of such examples is based on a modified version of Grice's second maxim of Quantity, which he calls the "I-Principle", largely equivalent to Horn's (Horn, 1984) "R principle":

Speaker's maxim: the maxim of Minimization. "Say as little as necessary" [...].

Recipient's corollary: the Enrichment Rule. Amplify the informational content of the speaker's utterance, by finding the most specific interpretation, up to what you judge to be the speaker's m-intended point [...].

Specifically:

a. Assume the richest temporal, causal and referential connections between described situations or events, consistent with what is taken for granted. (Levinson, 2000: 114)

Now an important aspect of Levinson's theory is that there are crucial differences between generalized and particularized implicatures: on his view, the former give rise to what he calls "utterance-type meanings" (2000: § 1.2), as opposed to the utterance-token meanings carried by particularized implicatures. But as the recipient's corollary makes clear, the process of enrichment is still construed as Gricean by him: the hearer has to find an interpretation that accords with what the speaker may have non-naturally meant. This is crucially different from the Fregean idea that the enrichment may be due to psychological associations that happen automatically in both the speaker and the hearer.

It is of course an empirical question which of these two types of explanation works best in which cases, but there seems to be empirical evidence that Frege was essentially right at least about some examples. In current psycholinguistics, there is the common assumption that text understanding requires readers to build so-called "situation models". Such models may involve things such as a map of the spatial relations between the objects that play a role in a text, but more importantly with respect to examples such as (4–1) or (4–2) they also involve categories such as motivation or causation (cf. Zwaan & Rapp, 2006: § 5; cf. Zwaan & Radvansky, 1998). Let us focus on motivation here. Frequently, fictional as well as non-fictional texts are essentially about a character trying to achieve a certain goal, and in such cases "the goals a character has for a particular situation are actively maintained in memory over the course of a reading experience" (Zwaan & Rapp, 2006: 742).

This may sound rather abstract, so consider an example. Poe's *The Cask of Amontillado* involves no more than two protagonists, Montresor and Fortunato, and the very first sentence makes Montresor's motives completely explicit: "The thousand injuries of Fortunato I had borne as I best could, but when he ventured upon insult, I vowed revenge." The story goes on to describe

how Montresor lures Fortunato into his wine cellar where he eventually buries him alive, immuring him behind a wall. Here are the final sentences of the story:

> I hastened to make an end of my labor. I forced the last stone into its position; I plastered it up. Against the new masonry I re-erected the old rampart of bones. For the half of a century no mortal has disturbed them. *In pace requiescat!*

It seems clear that a full understanding of this passage requires some kind of "enrichment". You will not understand them if you do not realize that Montresor executes the final steps of his revenge plan here. Now what distinguishes this case from the previous ones is that the explanatory mental state and the action that is explained by it are separated by more than 2000 words. Let us focus on the first and on one of the final sentences here:

> (4-4) [α] [W]hen [Fortunato] ventured upon insult, I vowed revenge. [≈ 2000 interjacent words] [ω] Against the new masonry I re-erected the old rampart of bones.

It seems clear that the action in [ω] is explained by [α], but at the same time the text does not encode this explanatory relation. So it must be "inferred" by a careful reader. But is it plausible to assume that a reader R attempts to find here "the most specific interpretation, up to what [R judges] to be the speaker's m-intended point" (Levinson, 2000: 114)? Such a kind of explanation may be plausible in simpler cases such as (4–3), but in the case of (4–4) there are two considerations that speak against such an explanation. *First*, the enrichment is non-local: the reader has to form a link between two sentences that occur, respectively, on the first and on the last page of the story. It is not clear whether the notion of implicature can handle such cases. In contrast, it is one of the insights of the situation model account that the reader's recognition of goals can "organize sequences of states and actions *over large surface distances in the text*" (Suh & Trabasso, 1993: 279, emphasis added). *Second*, the enrichment process in examples such as (4–3) involves nothing else than "replacing" a full stop by a conjunction:

> (4-3a) John fell and broke his leg [because] He [had] lost his grip on the cliff.

In contrast, understanding a story such as *The Cask of Amontillado* involves keeping track of various goals and their respective sub-goals. Accordingly, sentence [ω] would have to be enriched along the following lines:

> (4–4a) [ω] Against the new masonry I re-erected the old rampart of bones [which was the final step of immuring Fortunato, and I did so because I had vowed revenge].

It seems implausible that keeping track of several (hierarchically organized) goals while reading a story can be construed as an example of implicature. Note that the issue here is not whether the interpretation process requires conscious inferences: the neo-Gricean account and the psycholinguistic explanation are both non-committal as to how effortless these inferences are. (On Grice's view, implicatures are essentially *calculable*, but that does not necessarily exclude the existence of standardized implicatures that do not need to be *calculated*; cf. Bach & Harnish, 1979: 192–195). What is at issue here is rather the explanatory mechanism itself: for Griceans, understanding

enrichments essentially requires grasping "the speaker's m-intended point" (Levinson, 2000: 114) that goes beyond what was said. On the alternative view, a sentence describing a person's action will be read as (possibly) related to that person's motivational states, independently of what the author of the sentence may have meant with it. As said above, it is an empirical question whether a specific example is best explained in a psycholinguistic manner or, alternatively, via the Gricean mechanism. But the Fregean idea that *some* cases of coherence-maximizing enrichments are a matter of psychological associations seems to be on the right track.

Let us finally return to Blackburn's claim that stable ψ-effects may have much the same communicative effect as a linguistic convention. When characterizing such effects as communicative, it seems as though we need to rely on a non-Gricean notion of communication. But we shall see that things are more complicated. I think we ought to distinguish three kinds of cases here. The difference between these cases may be best explained by Frege's Napoleon example (repeated here):

> (4–1) Napoleon, [1] who recognized the danger to his right flank, [2] himself led his guards against the enemy position.
>
> (Frege, 1892: 47; FR 168)

If the side-thought that [2] happened because of [1] is a mere matter of psychological associations, then this would be an example of naïveté about ψ-effects: a speaker S may utter (4–1) because of having construed [1] as a reason for [2]; and an addressee H may simply hear [1] as offering a reason for [2]. In such a scenario, the side-thought will have been "transferred" although S may not have had the intention to convey it.

Next there are semantic delusions about ψ-effects—a phenomenon already noted by Frege: "And because [the side-thoughts] seem so naturally connected with our words, almost like the main thought itself, we then may also want to express such a side-thought." (Frege, 1897: 46, my translation). What Frege seems to be saying here is that a psychological association may create the illusion that a merely associated content is actually part of what we express when uttering a sentence. In such cases, a ψ-effect may become part of the intentional communicative content of an utterance even though our words do not actually express it. (If Blackburn's claim about "thick terms" is correct, then such terms may provide a neat example of such an illusion: living in a culture where sexual promiscuity, or the like, is generally considered morally objectionable may result in the impression that the word "lewd" carries an evaluation as part of its linguistic meaning, even though it does not.)

Finally, and I think most importantly, there is the manipulative use of ψ-effects. Such cases are similar to the ones Grice presented as counterexamples to his "first shot" at analysing the notion of meaning$_{NN}$. Consider one of his examples (and another example due to Frege):

> (4–5) [A leaves] B's handkerchief near the scene of a murder in order to induce the detective to believe that B was the murderer.
>
> (Grice, 1989: 218)

> (4–6) If a commander conceals his weakness from the enemy by making his troops keep changing their uniforms, he is not telling a lie; for he is not expressing any thoughts, although his actions are calculated to induce thoughts in others.
>
> (Frege, 1897: 152, 1997: 241)

In both of these cases, a person tricks another person into *believing* that something is the case by performing a *non-linguistic* action. But the phenomenon generalizes: the action may also be linguistic in character (for instance, using "cur" instead of "dog"), and the mental state that is induced may be something else than a belief (for example, a mental image or an emotion). What, exactly, is the difference between cases such as (4–5) or (4–6) and examples of communication proper? Grice's initial proposal for defining "communication", or "meaning$_{NN}$" created a huge debate, in which one counterexample after another led to more and more complicated analyses. But the contrast that is important here can be best brought out by Grice's first amendment:

Clearly we must at least add that, for *x* to have meant$_{NN}$ anything, not merely must it have been "uttered" with the intention of inducing a certain belief but also the utterer must have intended an "audience" to recognize the intention behind the utterance.

(Grice, 1989: 217)

This amendment is important here because it neatly distinguishes between ordinary cases of "open" communication and the manipulative employment of ψ-effects. Recall my brief discussion of code words and expressives from the previous sections. If words such as "cur" or "inner city" indeed have the power to induce all sorts of ψ-effects in an audience (stereotypical beliefs, mental images, and the like), then in order for these effects to actually take place it may be vital that the audience is unaware of the speaker's intention in using such words. This can be seen by the fact that making one's intention explicit in such cases may be self-undermining. Compare the following two utterances:

(4–7) The battle of Waterloo was fought in 1815, and I'm telling you that in order to inform you that this is what I believe.

(4–8) The cur howled the whole night, and I'm using the term "cur" in order to make you form a mental image of a scruffy dog.

Needless to say, both utterances are fairly odd. But there is a difference: in (4–7), the mundane intention of expressing one's belief by asserting something, which is perhaps constitutive of meaning$_{NN}$ something by a declarative sentence (cf. Grice, 1989: 110), is just made (overly) explicit. The cumbersome right conjunct can be understood as stressing the speaker's sincerity. The right conjunct of (4–8), in contrast, makes explicit how the word "cur" is supposed to act on the addressees' mind and thus effectively tells addressees that the speaker is attempting to manipulate them, as opposed to communicating with them.[19]

That manipulation must generally be latent in order to work efficiently is an insight that goes back to classical rhetoric, often summed up in the slogan "The art is to conceal the art" (*ars est celare artem*). The problems that may arise due to an all too obvious use of rhetorical figures are nicely brought out by (Pseudo-)Longinus:

There is an inevitable suspicion attaching to the sophisticated use of figures. It gives a suggestion of treachery, craft, fallacy [...]. So we find that a figure is always most effective when it conceals the very fact of its being a figure.

---

[19]Couldn't one say "The cur howled the whole night, and I'm using the term 'cur' (as opposed to 'dog') deliberately". Yes, but it depends on how one continues. Take a politically charged term such as "death tax" as an example. One might say "I'm using the term 'death tax' deliberately because it's the appropriate term". In contrast, "I'm using the term 'death tax' deliberately because I want you to acquire a negative attitude towards inheritance taxes" would be a self-undermining piece of propaganda. Thanks to an anonymous reviewer for pressing me on this issue.

(Longinus, 1995: 231, *On the Sublime*, 17,1)

What makes words such as "inner city" or "cur" (provided that they function roughly in the way I have assumed here) especially apt for manipulative purposes is precisely their inconspicuousness: in contrast to elaborate rhetorical figures that stand out as deviations from everyday language, they are part of our mundane vocabulary – unless of course they have been recognized as instruments of propaganda. (Much the same might be said about the non-poetic metaphors that abound in political speeches.)

5 | NON-ARBITRARY SIGNS

In the previous two sections, we have seen that ψ-effects may successfully explain social phenomena even though they are by their very nature subjective. Ψ-effects may offer a partial account of the social meaning of expressives or code words; and a successful explanation of how we read coherence relations into a text is perhaps best provided by psycholinguistics (and not by ordinary Gricean pragmatics). Let us now turn to another kind of case.

One of Frege's standard examples of the phenomenon he called "colouring" is the psychological effect induced by the mere sound quality of certain words.

> And in fact it cannot be denied that the spoken word affects the ideas we have just because it enters consciousness as a complex of auditory sensations. Right from the start we experience the series of sounds themselves, the tone of the voice, the intonation and rhythm with feelings of pleasure or displeasure. These sensations of sound are linked to auditory ideas that resemble them and these latter are linked in turn with further ideas reactivated by them. This is the domain of onomatopoeia.
>
> (Frege, 1897: 151, 1997: 239f.)

Frege's main aim in mentioning such issues was to elucidate his notion of thought by offering examples of phenomena that have nothing to do with the expression of thoughts, and the "pleasure or displeasure" we experience when hearing certain (chains of) words is of course a particularly apt example. At the same time, this may be a potential reason for downplaying the importance of such ψ-effects: one might argue that they are highly relevant to literary aesthetics, but simply irrelevant to the philosophy of language. We shall see, however, that Frege's observation that they "are linked in turn *with further ideas reactivated by them*" (emphasis added) speaks against ignoring them: in some cases, the sound quality of an item is not a mere matter of aesthetics, but has also to do with its emotive significance.[20]

Consider an example. In an experiment conducted by Myers-Schulz et al. (2013), English-speaking subjects were asked to match pictures with pseudo-words such as "bupaba" and "dugada". Among these pictures was an image of aggressively barking German Shepherd and an image of a cute St. Bernard puppy, and the subjects generally matched "bupaba" with the latter image, and "dugada" with the former. The prediction that "dugada" carries negative associations is based on some general claims about the association between certain phonetic features and

---

[20]Compare Frege (1882: 52, 1972: 86): "sounds have a more intimate relation to the emotions [*das Gemütsleben*] than shapes and colours do; and the human voice with its boundless flexibility is able to do justice to even the most delicate combinations and variations of feelings."

particular emotive effects. The details need not concern us here, but a particular prediction might be worth mentioning:

> Even in artistic contexts … these acoustic principles could be applied to evoke a particular emotional subtext. Indeed, our data suggest that "Darth Vader" … is an acoustically more appropriate name for an intergalactic miscreant than "Barth Faber".
>
> (Myers-Schulz et al., 2013: 7)[21]

According to Myers-Schulz et al., the "emotional subtext" carried by certain items would thus be non-arbitrary, and it seems plausible that this observation extends to the realm of "social meaning". Consider the word "rose" as an example. Famously, roses would smell as sweet if they were not called such. However, the social significance of roses, and the associations triggered by the word "rose", appear to depend, to some extent, on the fact that the word "rose" has an attractive sound quality, making it suitable for being used as a first name. Roses, arguably, would not be considered romantic if our ancestors had decided to call them, say, "pigsqueak".[22]

A particularly interesting example of items where the mere sound quality may carry emotional overtones are certain pejoratives. Quite a few slurs are phonetically similar to other "nasty words", and many of them are literally four-letter words: they are generally monosyllabic, have short vowels and voiceless consonants, mostly fricatives and plosives. These features in combination make them sound ugly and aggressive, and often even small changes, such as replacing a short vowel with a long one, would suffice to mitigate their derogatory sound quality.[23]

Note that this is not to be understood as yet another "deflationist" theory of slurs. I am not claiming here that slurs are nothing else than aggressive-sounding words for a particular group. Not all of the items that are standardly categorized as slurs conform to the above-mentioned sound pattern, and I obviously do not wish to claim that the ugliness of certain slurs may offer a full explanation of what is going on when speakers are using such words.

Quite to the contrary, I think that the present debate on slurs is often marred by two problematic assumptions.[24] Scholars frequently assume, first, that slurs form a homogeneous category and, second, that we ought to identify a single mechanism (be it semantic or pragmatic) that fully explains how these items work. I think we ought to reject both of these assumptions. There are, first, crucial differences between such items: for instance, a word like "kraut", as being used by an educated UK speaker in the early 1990s, may "emphasize a sense of difference" (Blackburn, 1992: 294), but in contrast to many other "slurs" it is clearly neither derogatory nor a term that may be construed as expressing a negative attitude. Second, it seems plausible that *fully* understanding some of these items may require taking *several*

---

[21]And neither would "Hubert Farnsworth" be an apt name for a supervillain. See also Köhler's (Köhler, 1947: 224–5) observation on how nonsense words such as "takete" and "maluma" are associated with certain forms—a phenomenon now generally known as "bouba/kiki effect".

[22]Similar things might be said about the graphical qualities of written words. Compare Wittgenstein's (1986, § 167) remark on spelling (reforms): "Think of the uneasiness we feel when the spelling of a word is changed. (And of the still stronger feelings that questions about the spelling of words have aroused.)"

[23]For recent work on the sound symbolism of swear words, compare Lev-Ari and McKay (2023). Mandelbaum and Young (forthcoming) reports on an experimental study on the "sound of slurs".

[24]Compare Stanley (2015: 150f) who similarly argues against the assumption that slurs are "special".

mechanisms into account. Anderson and Lepore have argued that an account of slurs in terms of Fregean tone, i.e. a purely psycholinguistic account, "can't be the whole story" (2013: 33). I agree with them, but something that is not the whole story can still be an *essential part* of the whole story.

6 | THE ESSENTIALIZING EFFECT OF KIND-DENOTING TERMS

Let us finally turn to a kind of ψ-effect that is not discussed by Frege. Consider the following two sentences:

> (6–1) Lorraine is blonde.

> (6–2) Lorraine is a blonde.
>
> (Ritchie, 2021: 571)

These two sentences are, arguably, truth-conditionally equivalent. However (6–2), involving not an adjective but a predicate nominal, seems to invite certain inferences: (6–2) somehow suggests that being (a) blonde is an essential property of Lorraine's, which in turn may explain some of her other properties. This intuition that certain nouns have an essentializing effect is clearly supported by experimental research (cf. Ritchie, 2021: 573–4 for an overview). For instance, when children are given sequences of sentences like

> (6–3) Rose is 8 years old. Rose eats a lot of carrots. {[a] She is a carrot-eater. / [b] She eats carrots whenever she can.}
>
> (Gelman, 2004: 407)

and are then being asked questions such as whether Rose will eat lots of carrots when she is grown up, they are more likely to answer such questions positively when the sequence involves an [a]-type sentence, i.e. a noun. According to Gelman (2004: 408), "nouns imply that a category is relatively more stable and consistent over time and contexts than adjectives or verbal phrases", and this raises the question of what is meant by "imply" here.

In a recent paper, Ritchie (2021: 586) has claimed that "kind-denoting terms trigger a presupposition that there is ... an inductively potent kind with stable and explanatory membership", and the term "presupposition" here is not loose talk. On her view, the following two cases are *exactly* parallel (where ">>" denotes the relation of presupposition):

> (6–4) He who discovered the elliptic form of the planetary orbits died in misery. [>> There is a unique person who discovered the elliptic form of the planetary orbits]
>
> (Frege, 1892: 39)

> (6–5) Anna is a female. [>> There is an inductively potent kind, females, with stable and explanatory membership.]
>
> (Ritchie, 2021: 585)

A leitmotif running through this paper is that some phenomena that scholars have accounted for in terms of semantics or pragmatics are better explained (in part) by ψ-effects, and I think that (6–5) is yet another example. Although I would concede that this a complicated case, there seem to be at least four points that speak against construing such nouns as presupposition triggers. *First*, whatever exactly is conveyed by (6–5), it seems to be much less determinate than the alleged presupposition of that sentence, especially when what is presupposed is rendered in what I take to be an overintellectualized way. The fact that this content is rather elusive can more plausibly be explained by ψ-effects. *Second*, the assumption that "a female" carries a presupposition proper does not seem to fit its radical projection behaviour. "A female" appears to suggest an "explanatory membership" relation (or the like) regardless of how that term is embedded. Typical presupposition triggers, in contrast, can easily be neutralized: neither "Macron isn't the King of France" nor "If France is still a Monarchy, then the King of France must be rich" commits the speaker to the existence of a French monarch.[25] *Third*, recall what I said about code words in Section 2: one cannot challenge the suggestion carried by "welfare" by a "wait a minute"-style of reply, and similarly one cannot felicitously react to (6–5) by saying "Hey, wait a minute, there isn't an inductively potent kind …". The suggestion carried by "a female" seem to be much more "backgrounded" than is the case with ordinary non-at-issue contents. *Finally*, the experimental evidence merely indicates that subjects are *more likely* to make essentializing assumptions when being given sentences involving nouns such as "carrot-eater". The existence presupposition carried by descriptions, in contrast, is not a matter of statistics: if sentences involving descriptions do trigger such presuppositions (which is not always the case), they reliably commit the speaker to the existence of a certain entity.

7 | CONCLUSION: TAXONOMIZING Ψ-EFFECTS

In the last section, I conceded that nouns such as "a female" are a complicated issue, but let us assume that I am on the right track here. What makes this example particularly interesting on that assumption is the fact that ψ-effects may also be triggered by the grammatical properties of a term (noun vs adjective), whereas in the previous examples they were carried either by the meaning or by the sound quality of an item. This also suggests a neat way of taxonomizing ψ-effects. On my view, we ought to ask two questions:

(Q1) What triggers a ψ-effect?

(Q2) What kind of mental state is triggered?

In the course of this paper, we have encountered three answers to (Q1): there are phonetic, semantic, and grammatical effects. Listing the possible answers to (Q2) is somewhat more complicated since there is no uncontroversial classification of mental states. But for present purposes, it might be useful to roughly distinguish four kinds of effect: effects may be aesthetic experiences, emotions, cognitive states, and mental images.

---

[25]This observation suggests that a broadly semantic explanation of the contribution of phrases such as "a female" might work better if one construes them as Pottsian "conventional implicatures" (CIs). On Potts's (Potts, 2005: 34–36) view, CIs differ from presuppositions precisely in being "scopeless".

The examples I have discussed in this paper may then be represented in the following table:

	Phonetic	Grammatical	Semantic
Aesthetic	"rose", rhyme and rhythm		
Emotive	"Darth Vader", some slurs		"coffin", "welfare", "cur"
Cognitive		"a female"	"welfare", textual coherence
Mental Images	German "Jedem das seine"		"cur", "welfare"

I have repeatedly stressed that the exact nature of a ψ-effect is an empirical issue. So the above table is not to be understood as the final wisdom. At the same time, the distinction between kinds of triggers involves several complications. For instance, I have listed the German example from Section 2 as a phonetic effect, but the effect is arguable not totally independent of the meaning of that phrase. Similarly, the specific contribution of "welfare" may be different from the contribution of synonymous terms, which would mean that the effect is triggered by a concept under a phonetic guise.

However, the aim of this paper has not been to argue for a particular account of, say, code words. Instead, I have attempted to show that we generally ought to resist the tendency to explain certain linguistic phenomena exclusively in terms of either semantics or pragmatics. We should not ignore the obvious and introspectively available fact that words have all sorts of effects on what is going on in our minds, and if I am right such ψ-effects are crucial for fully understanding the overall communicative import of some of our utterances.[26]

ORCID
Thorsten Sander https://orcid.org/0000-0002-0880-8843

REFERENCES

Anderson, L., & Lepore, E. (2013). Slurring words. *Noûs*, *47*(1), 25–48.

Bach, K., & Harnish, R. M. (1979). *Linguistic communication and speech acts*. MIT Press.

Beaver, D., & Stanley, J. (2019). Toward a non-ideal philosophy of language. *Graduate Faculty Philosophy Journal*, *39*(2), 503–547.

Blackburn, S. (1992). Through thick and thin. *Proceedings of the Aristotelian Society Supplementary*, *66*, 284–299.

Cappelen, H. (2018). *Fixing language: An essay on conceptual engineering*. Oxford University Press.

Cappelen, H., & Dever, J. (2019). *Bad language*. Oxford University Press.

Carston, R. (2018). Figurative language, mental imagery, and pragmatics. *Metaphor and Symbol*, *33*(3), 198–217.

Devitt, M., & Sterelny, K. (1999). *Language and reality: An introduction to the philosophy of language* (2nd ed.). Blackwell.

Dummett, M. (1981). *Frege: Philosophy of language* (2nd ed.). Harvard University Press.

Foolen, A. (2015). Word valence and its effects. In U. Lüdtke (Ed.), *Emotion in language* (pp. 241–255). John Benjamins.

Frege, G. (1882). Ueber die wissenschaftliche Berechtigung einer Begriffsschrift. *Zeitschrift für Philosophie und philosophische Kritik*, *81*, 48–56.

Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, *100*, 25–50.

---

[26]I would like to thank an anonymous referee for several helpful comments and suggestions.

Frege, G. (1897). Logik. In H. Hermes, F. Kambartel, & F. Kaulbach (Eds.), *Gottlob Frege: Nachgelassene Schriften* (pp. 137–163). Felix Meiner.

Frege, G. (1918/1919). Der Gedanke. Eine logische Untersuchung. *Beiträge zur Philosophie des deutschen Idealismus*, *2*, 58–77.

Frege, G. (1972). In T. W. Bynum & T. T. W. Bynum (Eds.), *Conceptual notation, and related articles*. Clarendon Press.

Frege, G. (1979). In P. Long, R. White, T. H. Hermes, F. Kambartel, & F. Kaulbach (Eds.), *Posthumous writings*. Basil Blackwell.

Frege, G. (1997). In M. Beaney (Ed.), *The Frege reader*. Blackwell.

Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences*, *8*(9), 404–409.

Grice, H. P. (1989). *Studies in the way of words*. Harvard University Press.

Haslanger, S. (2013). Social Meaning and Philosophical Method. American Philosophical Association 110th Eastern Division Annual Meeting (December 2013). http://hdl.handle.net/1721.1/97049

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1992). *Emotional contagion*. Cambridge University Press.

Hobbs, J. R. (1979). Coherence and coreference. *Cognitive Science*, *3*(1), 67–90.

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, form, and use in context: Linguistic applications* (pp. 11–42). Georgetown UP.

Horn, L. R. (2007). Toward a Fregean Pragmatics: Voraussetzung, Nebengedanke, Andeutung. In L. R. Horn & I. Kecskes (Eds.), *Explorations in pragmatics* (pp. 39–69). de Gruyter.

Horn, L. R. (2013). I love me some datives: Expressive meaning, free datives, and F-implicature. In D. Gutzmann & H.-M. Gärtner (Eds.), *Beyond Expressives: Explorations in use-conditional meaning* (pp. 151–199). Brill.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Strauss and Giroux.

Kant, I. (1991). In M. Gregor (Ed.), *The metaphysics of morals*. Cambridge University Press.

Khoo, J. (2017). Code words in political discourse. *Philosophical Topics*, *45*(2), 33–64.

Klemperer, V. (2013). *The language of the third Reich* (M. Brady, Trans.). Bloomsbury.

Köhler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology*. Liveright.

Lepore, E., & Stone, M. (2018). Slurs and tone. In A. Coliva, P. Leonardi, & S. Moruzzi (Eds.), *Eva Picardi on language, analysis and history* (pp. 205–217). Springer.

Lev-Ari, S., & McKay, R. (2023). The sound of swearing: Are there universal patterns in profanity? *Psychonomic Bulletin & Review*, *30*, 1103–1114.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press.

Longinus. (1995). On the sublime (W.H. Fyfe & D. Russell, Trans.). In *Aristotle, Poetics* (pp. 143–307). Harvard University Press.

Ludlow, P. (2014). *Living words: Meaning underdetermination and the dynamic lexicon*. Oxford University Press.

Mandelbaum, E., & Young, S. (forthcoming). The sound of slurs: Bad sounds for bad words. *Oxford Studies in Experimental Philosophy*. https://philpapers.org/rec/MANTSO-26

Myers-Schulz, B., Pujara, M., Wolf, R. C., & Koenigs, M. (2013). Inherent emotional quality of human speech sounds. *Cognition & Emotion*, *27*(6), 1105–1113.

Picardi, E. (2007). On sense, tone and accompanying thoughts. In R. E. Auxier & L. E. Hahn (Eds.), *The philosophy of Michael Dummett* (pp. 491–520). Open Court.

Pinker, S. (2014). *The sense of style: The thinking person's guide to writing in the 21st century*. Penguin Books.

Potts, C. (2005). *The logic of conventional implicatures*. Oxford University Press.

Potts, C. (2015). Presupposition and Implicature. In S. Lappin & C. Fox (Eds.), *The handbook of contemporary semantic theory* (2nd ed., pp. 168–202). Wiley Blackwell.

Quintilian. (1854). In E. Bonnell (Ed.), *Institutionis oratoriae libri duodecim, vol. 1*. Teubner.

Ritchie, K. (2021). Essentializing Inferences. *Mind & Language*, *36*, 570–591.

Salmon, W. (2011). Conventional implicature, presupposition, and the meaning of *must*. *Journal of Pragmatics*, *43*(14), 3416–3430.

Sander, T. (2019). Two misconstruals of Frege's theory of colouring. *The Philosophical Quarterly*, *69*(275), 374–392.

Sander, T. (2021). Fregean side-thoughts. *Australasian Journal of Philosophy*, *99*(3), 455–471.

Saul, J. (2018). Dogwhistles, political manipulation, and philosophy of language. In D. Fogal, D. W. Harris, & M. Moss (Eds.), *New work on speech acts* (pp. 360–383). Oxford University Press.

Saul, J. M. (2002). Speaker meaning, what is said, and what is implicated. *Noûs*, *36*(2), 228–248.

Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy*, *25*(5/6), 701–721.

Stalnaker, R. (2014). *Context*. Oxford University Press.

Stanley, J. (2015). *How propaganda works*. Princeton University Press.

Stevenson, C. L. (1944). *Ethics and language*. Yale University Press.

Stojnić, U., & Lepore, E. (2022). Inescapable articulations: Vessels of lexical effects. *Noûs*, *56*(3), 742–760.

Suh, S., & Trabasso, T. (1993). Inferences during reading: Converging evidence from discourse analysis, talk-aloud protocols, and recognition priming. *Journal of Memory and Language*, *32*(3), 279–300.

Tonhauser, J., Beaver, D., Roberts, C., & Simons, M. (2013). Toward a taxonomy of projective content. *Language*, *89*(1), 66–109.

Väyrynen, P. (2013). *The lewd, the rude and the nasty: A study of thick concepts in ethics*. Oxford University Press.

White, I. K. (2007). When race matters and when it doesn't: Racial group differences in response to racial cues. *American Political Science Review*, *101*(2), 339–354.

Williamson, T. (2009). Reference, inference and the semantics of pejoratives. In J. Almog & P. Leonardi (Eds.), *The philosophy of David Kaplan* (pp. 137–158). Oxford University Press.

Wilson, D., & Carston, R. (2019). Pragmatics and the challenge of 'non-propositional' effects. *Journal of Pragmatics*, *145*, 31–38.

Wittgenstein, L. (1986). *Philosophical investigations* (G. E. M. Anscombe, Trans. 3. ed.). Basil Blackwell.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162–185.

Zwaan, R. A., & Rapp, D. N. (2006). Discourse comprehension. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 725–764). Elsevier.