Speaker's Reference, Semantic Reference, Sneaky Reference

Eliot Michaelson

King's College London

Abstract

According to what is perhaps the dominant picture of reference, what
a referential term refers to in a context is determined by what the
speaker intends for her audience to identify as the referent. I ar-
gue that this sort of broadly Gricean view entails, counterintuitively,
that it is impossible to knowingly use referential terms in ways that
one expects or intends to be misunderstood. After exploring several
possible responses—each of which ultimately proves unsatisfactory—I
sketch an alternative which can better account for such opaque uses
of language, or what I call 'sneaky reference'. I close by reflecting on
the ramifications of these arguments for the theory of meaning more
broadly, as opposed to just the theory of reference.

1 Introduction

Speakers sometimes say one thing in order to get their audience to believe
something else. A politician might, for instance, implicate or merely hint
at something without actually saying it.[1] This much is commonly accepted,
and a great deal of work has already gone into trying to understand the
mechanisms whereby such implicating and hinting takes place. But there is
another way of getting one's listener to believe something without actually

---

Forthcoming in *Mind & Language.*

[1]E.g. Trump's "If she gets to pick her judges, nothing you can do, folks. Although the
Second Amendment people—maybe there is, I don't know" (`http://nyti.ms/2ayrB9J`).

saying that thing, one which has gone largely unnoticed in the philosophical literature to date: deliberately saying something that one intends for one's audience to misunderstand. Interestingly, the existence of this phenomenon will prove to be in tension with a wide range of contemporary theories of meaning. This essay aims (i) to explore why this tension arises, and (ii) to offer an alternative which is both independently plausible and capable of accounting for the possibility of such opaque speech.

To get clearer on the phenomenon in question, it will help to get an example on the table straightaway. To that end, let us turn to one of the few places in the philosophical corpus where cases of deliberately speaking so as to be misunderstood have actually been mentioned, Lewis (1969, p. 193). Lewis' main concern at this point in his essay was to introduce a notion he called 'minimal truthfulness'—the analogue of his 'truthfulness' for languages that include homophony or other forms of ambiguity. To hammer home how little is required to count as minimally truthful, Lewis tells us that even '[a] trickster is being truthful in this minimal way if, knowing that Owen is going to the shore of the river, he says "Owen is going to the bank" during a conversation about Owen's lack of cash' (loc. cit.).[2] Viewing ambiguity as a relatively minor issue, however, Lewis quickly moves on from this case.

Perhaps he should have lingered. Though apparently unnoticed by Lewis, the case actually runs counter to an extremely widespread assumption about meaning: that to mean something with one's words is ultimately a matter of manifesting a particular sort of intention in uttering those words, an intention that one's listener come to believe (or to entertain, or to believe that one believes, etc.) that very same something partly in virtue of recognizing this very intention. Due to its origins in Grice (1957), I will call this the 'core Gricean thesis'.[3] The popularity of this thesis extends well beyond self-identified Griceans, however.[4] For, unlike early Lewis, most theorists nowadays take ambiguity and other forms of underdetermination to be pervasive in human language. And the thesis that what disambiguates an utterance of an ambiguous sentence is the fact that the speaker means something more specific—in roughly Grice's sense—in uttering that sentence

---

[2]For similar cases, set to unrelated argumentative purposes, see Kotatko (1998) and Weatherson (2002); such cases are also not at all uncommon in literature.

[3]For various refinements of the core thesis, see *inter alia* Searle (1969), Schiffer (1972), Bennett (1976), Vlach (1981), Avramides (1989), and Neale (1992).

[4]For theories of meaning not typically considered Gricean, but still committed to this core Gricean thesis, see Davidson (1986) and Brandom (1994).

has proven to be a simple and elegant way of explaining our judgments on truth and falsity in a wide range of cases.[5] The exception, of course, is Lewis' trickster case and others like it.

My first task in this essay will be to assess whether a trickster-type objection to the core Gricean thesis can in fact be sustained, or whether there is some way around it. My second task will be to sketch a viable alternative to the Gricean theory. Towards both these ends, it will prove helpful to restrict our attention to a particular class of terms, preferably one that has been well-explored by Griceans. Below, I will focus in on singular referential terms like names and demonstratives. Not only is this familiar territory for Grieans, but many have claimed that these terms are particularly amenable to a Gricean analysis.[6] In restricting our attention in this way, we are thus doing the Gricean no disservice.

Here is the plan for what follows: in §2, I introduce the Gricean theory in more detail and explain how it has been applied to singular referential terms in particular. In §3, I offer several cases of what I will call 'sneaky reference'— essentially, referential analogues of Lewis' trickster—and explore why Gricean theories of meaning look to be ill-situated to explain these cases. §4 explores three potential responses on behalf of the Gricean. In §5, I introduce my alternative Direct Expression View of meaning and reference, according to which meaning is to be understood in terms of *voicing one's thoughts* or *putting one's thoughts into words*. §6 responds to some objections, and §7 concludes with some broader lessons about the nature of meaning.

2   The Gricean Theory of Reference

We begin with Grice (1957)'s core notion of *speaker's meaning*:

> "*A* meant something by x" is (roughly) equivalent to "*A* intended the utterance of x to produce some effect in an audience by means of the recognition of this intention."[7] (p. 385)

The crucial point for us is that Grice conceives of the act of meaning something as requiring the speaker's having a particular sort of audience-directed

---

[5]Lewis himself subsequently endorsed a thesis along these lines in his (1979).

[6]Cf. Stine (1978), Schiffer (1981), Bach (1992), Neale (2004), and Buchanan (2014).

[7]I have elided Grice's 'NN' subscript on 'meant', which indicated that he was talking about 'non-natural' meaning. I will take this as a given throughout.

intention. Such intentions are: (i) reflexive, so they are meant to be recognized and fulfilled partly on the basis of that recognition; and (ii) involve an intended effect which somehow encodes the content meant. It is this latter aspect of the Gricean program that will be called into question below.[8]

Restricting our attention to utterances of declaratives, we can get (ii) more clearly in view. For Grice, the intended effect of a declarative is: to bring it about that one's audience comes to believe p. So, at least for declaratives, the content of this intended belief, i.e. p, reveals the content meant by the speaker in uttering the relevant declarative sentence, i.e. p.

Subsequent authors have argued that belief is too strong and have instead suggested that the relevant effect should be 'entertaining p', 'believing that the speaker believes that p', etc.[9] For our purposes, we can afford to be ecumenical. Thus, we can restate the core thesis as:

> **MEAN**   To *mean that p* by an utterance U is to intend for one's audience to come to believe that p (or to entertain p, or to believe that one believes p, etc.) on the basis of hearing U, partly in virtue of their recognizing this very intention.

Note that this definition will apply equally well to both literal and non-literal sorts of meaning, like conversational implicatures. Since we'll really only be interested in the former, we need some way of restricting this thesis.

Grice (1969) provides just such a way in the course of discussing what he calls meaning something 'by the *words which [one] uttered*' (p. 162). For brevity's sake, I'll call this 'literal meaning':

> **MEAN$_L$**   To *literally mean that p* by an utterance U is to intend for one's audience to come to believe that p (or to entertain p, or to believe that one believes p, etc.) on the basis of hearing U, partly in virtue of their recognizing this very intention, and where p is among the standard meanings of the utterance-type of which U is a token.

The key difference between speaker's meaning and literal meaning is that the latter is required to conform to the standard usage of the relevant utterance-type. Grice calls this sort of standard *timeless meaning*, though we might

---

[8]This differentiates the present critique from those of Laurence (1998) and Glüer & Pagin (2003). These authors target the reflexivity of Gricean meaning rather than content-matching. But one could still endorse (ii) while rejecting (i).

[9]See Vlach (1981) and Neale (1992) for helpful overviews.

equally well think of it as a sort of 'conventional' or 'linguistic' meaning. What matters is just that this type of meaning will typically specify a range of ways that an utterance-type might be aptly used on a given occasion—to mean something true—as opposed to just a single way it can be used.[10]

Lastly, we need a way of talking about the meanings of utterances themselves as opposed to speakers' acts of meaning things with those utterances. Grice (1969) calls this *applied timeless meaning* and specifies that this sort of meaning 'gives one the correct reading of the utterance on a given occasion' (p. 148). Later, he introduces the category of *what is said, strictly speaking*—which appears to serve much the same role, but which makes the contrast with what is implicated, and thus never explicitly said, more apparent (Grice 1989, p. 25).[11] Adopting this latter terminology, we can put the idea as:

> **SAY**   What is said, strictly speaking, by an utterance of U is whatever the speaker literally means by her utterance of U.

To be clear, this notion of what is said isn't meant to capture everything we might naively judge to be 'said' by an utterance, but rather only that which is relevant to the strict truth or falsity of the utterance itself. This notion thus looks rather similar to what other philosophers have called the 'assertoric', 'semantic', or 'truth-conditional' content of an utterance.

The claim that the strict truth conditions of an utterance depend on speaker's meaning, even literal speaker's meaning, is more controversial than the claim that **MEAN** captures a core truth about meaning. When it comes to demonstratives and other highly context-sensitive terms, however, something along the lines of **SAY** is very commonly assumed. To see why, let us turn very briefly to the theory of demonstratives.

According to Kaplan (1989), the starting point for most contemporary work on demonstratives, demonstrative reference depends on the speaker's intentions. More specifically, Kaplan claims that an utterance of a demonstrative like 'this' or 'that' refers, in context, to whatever the speaker intends for it to refer to (Ibid., pp. 587–88).[12] In its favor, the view is admirably sim-

---

[10]**MEAN**$_L$ is also rather close to Searle (1969)'s preferred notion of meaning, which involves intentions for one's utterance to be interpreted in a certain way (p. 48).

[11]Grice himself tied these two notions together in the preface he added to 'Utterer's Meaning and Intentions' when it was reprinted in *Studies in the Way of Words*.

[12]Actually, Kaplan (1989) restricts this claim to just 'perceptual uses' of demonstratives, whatever exactly those are (p. 582). For simplicity's sake, I'll stick to the unrestricted analogue of Kaplan's actual view—which is how it is usually presented.

ple and respects the flexibility of these terms. What's more, unlike the view that demonstrative reference is fixed by an accompanying ostensive gesture, Kaplan's 'intentionalist' theory can account for the fact that demonstratives can refer even in the absence of any such gesture.[13]

The simplicity of the view comes at an apparent cost, however. On Kaplan's theory, it would appear that one can use *any* demonstrative one likes, in *any* context whatsoever, to refer to whatever one likes. The problem is that Kaplan leaves the notion of *intending to refer* unanalyzed, and hence apparently unconstrained. Without some constraints on these intentions, the Kaplanian theory looks bound to over-generate predictions of referential success. For even if demonstratives are highly flexible terms, they are not infinitely flexible. Kaplan's theory, in other words, risks becoming a 'Humpty Dumpty theory' in the sense that it allows speakers to make such terms refer willy-nilly—something which seems beyond their actual power to achieve.[14]

A clear illustration of the potential problem here can be found in Reimer (1991, p. 182). Reimer asks us to consider a case where the speaker is touching a dog, Fido, and starts to say:

(1)   This is Fido.

Just as she begins, however, Fido bolts and is replaced by Spot.[15] According to Reimer, 'the speaker's utterance was false, and was so on account of the fact that the speaker said falsely of Spot that he was Fido' (Ibid., p. 182). The Kaplanian, in contrast, looks set to predict the opposite.

Responding to this sort of worry, Bach (1992) proposes to couple Kaplan's theory with a Gricean understanding of the relevant intentions. While Reimer's speaker might well have intended to refer to Fido, Bach reasons that she did so only derivatively. Her specifically communicative intention, in contrast, was for her listener to recognize the object she was touching as the referent—for it is only this latter intention that the speaker can reasonably intend for her listener to recognize directly on the basis of her utterance. The speaker's mistake was to think that this object would remain Fido for the duration of her utterance, and thus that, by recognizing her communicative intention, the speaker would *de facto* recognize Fido as the referent.

---

[13]For the gesture theory, see McGinn (1981).

[14]Cf. Reimer (1991, 1992). The 'Humpty Dumpty' terminology comes from MacKay (1968).

[15]Reimer actually uses a complex demonstrative; I have simplified the case to avoid irrelevant complications.

Relying on Gricean communicative intentions to fix reference, Bach claims, thus yields the right result in Reimer's case: the speaker's utterance of 'this' refers to Spot rather than to Fido in virtue of Spot's being the target of the speaker's genuinely communicative intention. In slogan form: only *communicative* referential intentions serve to fix reference (Ibid., p. 143–4).[16]

The appeal of Bach's style of reasoning extends well beyond demonstratives, to other sorts of referential terms as well, e.g. names and pronouns.[17] Thus, we might introduce an analogue of literal meaning, literal reference:

> **REFER**$_L$ To *literally refer to O* by an utterance of a referring term e is to intend for one's audience to identify O as the referent (or as the speaker's intended referent, etc.) on the basis of hearing e, partly in virtue of their recognizing this very intention and where O is among the things to which e can standardly and felicitously be used to refer in the relevant context.

Functionally, literal reference partially determines the literal meaning of the overall utterance, and thereby contributes to what is said, strictly speaking, by means of that utterance. If one literally refers to O with some expression e in the course of uttering U, then we should expect for the strict truth and falsity of U to depend, at least in part, on how things are with O. Doing so, according to **REFER**$_L$, requires not just an intention for one's audience to recognize O as the referent, but also that the use of e to refer to O on this occasion comport with e's timeless meaning.[18]

Below, it will be important to distinguish between literal meaning and literal reference, on the one hand, and the particular *theories* of each that I have just sketched. Literal meaning and literal reference are functional concepts, best understood in terms of the role each plays in a broader explanatory project. For declaratives at least, literal meaning is marked out by its giving the strict truth conditions of an utterance. Literal reference, in turn, serves to make these truth conditions object-dependent. What Grice offers is a particular theory, namely **MEAN**$_L$, of what makes some content

---

[16]See Stine (1978) and Schiffer (1981) for earlier versions of this sort of view. More recent versions include Neale (2004), Stokke (2010), Buchanan (2014), and King (2014).

[17]On names, see Grice (1968, esp. pp. 237–38) and, more recently, Buchanan (2014).

[18]So, for example, while there is undoubtedly some sense in which a name like 'Fido' can be used to refer to Spot, and even to successfully communicate something about Spot, such a use will not comport with the timeless meaning of 'Fido'. 'Fido', I presume, is standardly used to refer to things named 'Fido', not to things named 'Spot'.

realize the literal meaning role for an utterance. Bach then extends this line of thought to generate a theory, namely **REFER**$_L$, of what makes a particular object realize literal reference role for a singular referential term. It is the combination of these views that I will call the 'Gricean Theory of Reference'.

Crucially, other accounts of literal meaning and literal reference are possible—including other intentionalist accounts. Indeed, the ultimate aim of this inquiry will be to offer an intentionalist account of literal meaning and literal reference that is capable of avoiding the Humpty Dumpty worry while nonetheless denying the core Gricean thesis that the sorts of intentions relevant to fixing meaning and reference are, ultimately, audience-directed.

One final clarification before moving on to the argument: in his (1977), Kripke famously distinguishes between what he calls 'speaker's reference' and 'semantic reference'. For Kripke, semantic reference, like our literal reference, helps to determine the strict truth or falsity of an utterance. But Kripke also assumes a tight connection between semantic reference and convention; in the case of names, for instance, he assumes that their reference, in context, is determined entirely by convention (Ibid., pp. 263–4). Since this sort of picture looks rather implausible in the case of demonstratives—and, indeed, even in the case of names with multiple bearers—I do not want to assume any such tight connection between strict truth and convention.[19] So it will be preferable to stick with our notion of literal reference instead.

3   Sneaky Reference

Having spelled out the details of the Gricean Theory of Reference above, I will now argue that there are cases this theory is ill-suited to capture. As advertised at the outset, these cases will be referential analogues of Lewis' trickster case—cases in which the apparent literal referent is one that the speaker neither intends nor expects for the listener to recover.[20]

First, consider a case involving demonstratives. Suppose that you and I both work at a large investment bank. One day, we run into each other

---

[19]Compare Bach (2005), who accepts that reference helps to determine utterance-truth and -falsity, but who infers from the non-conventionality of reference that truth conditions must be a pragmatic phenomenon rather than a semantic one.

[20]Weatherson (2002) describes some similar cases, also involving reference—though indexical rather than demonstrative reference. This difference matters in the present context, since even some Griceans (e.g. Bach 2005) have been apt to treat indexical reference as a special case, one that is not adequately captured by the Gricean Theory.

near a window overlooking the parking lot. You ask me what I did with the obscenely large bonus we were just paid. Being a bit of a prankster, I point out the window towards a small cluster of cars and tell you:

(2)   I bought that.

The cluster at which I am pointing contains several nondescript automobiles and, in addition, a McLaren F1. As it happens, I spent a small part of my bonus on the lightly-used Toyota Camry parked next to the McLaren.

Suppose further that I fully understand how you are likely to respond here: by taking me to be talking about the McLaren. Still, it seems possible for me to intend—at least in Kaplan's thin sense rather than in the thicker Gricean sense—for my utterance of 'that' to refer to the Camry. In other words, it seems possible for me to intend only to trick you, not to lie to you. Plausibly, when I do this, I manage to say something minimally true. My utterance is true, at least in Lewis' minimal sense, but highly misleading.

Here is a second example, this one involving names.[21] Suppose that you and I regularly eat lunch together. Today though, I ate lunch early. So when you stop by my office to invite me to lunch, I say:

(3)   Sorry, I've already eaten. But Sam is having lunch at the cafeteria.

Suppose further that there are actually two Sams in our department—one of whom you rather like (call her 'Sam the Good', or 'Sam_G' for short) and one of whom you find rather annoying (call her 'Sam the Mildly Annoying', or 'Sam_M' for short). As it happens, it is Sam_M, not Sam_G, who is eating lunch at the cafeteria. I know this, and I also know that you are overwhelmingly likely to take me to be talking about Sam_G. In fact, that's exactly what I intend for you to do. My hope is that, on seeing Sam_M eating lunch alone, you will be moved to join her. Then hopefully you will come to realize that, really, Sam_M isn't so bad. To my mind, this would be an excellent outcome, since then I can start inviting both of you to my weekly Dungeons & Dragons game. At present, you keep refusing to come if Sam_M will be there playing her surly, foul-mouthed, half-orc rogue.[22]

---

[21]For present purposes, I'm assuming that names are referential terms. While this assumption is widespread, it is by no means universal (cf. Graff Fara 2015).

[22]Note that on Kaplan (1990)'s version of Millianism, these two Sams won't actually share a name. Still, it remains for Kaplan to specify which name (effectively, which of our subscripted names) was really uttered in this context. A brute appeal to causal chains won't do, since two competing causal chains clearly helped to generate this utterance.

Again assuming that I was keen only to trick you, not to lie to you outright, it strikes me that my utterance of (3) was true in these circumstances—true, but highly misleading.[23]

In each case, the Gricean Theory of Reference predicts the opposite. By **REFER**$_L$, it predicts that the speaker literally refers to the McLaren and to Sam$_G$ with her utterance of 'that' and 'Sam', respectively. By **SAY**, the strict truth or falsity of each utterance depends on how things are with each of these objects. Intuitively, these would seem to be the wrong results.

To reinforce the point, consider whether either of our speakers has lied. Suppose that one were to accuse the speaker of lying in the two-Sams case. She might justly reply: 'I didn't lie to you. I told you that Sam was having lunch at the cafeteria and Sam *was* having lunch at the cafeteria!' If that's right, then we have reason to think that the speaker didn't say something about Sam$_G$ here: if she had, then she would have said p, strictly speaking, while both believing that $\neg p$ and intending to deceive her listener about the truth of p. While there is significant disagreement regarding the exact nature of lying, there is near universal agreement that these three conditions are jointly sufficient.[24] So either we must reject the claim that what is said, strictly speaking, is relevant to determining whether someone has lied, or else we must reject the Gricean theory of what is said, strictly speaking.[25]

It is worth pointing out that very little hinges on the particulars of these examples. The basic claim is schematic: take any sentence S that contains either a context-sensitive or ambiguous element, such that it can be interpreted as either S_1 or S_2. Now imagine a speaker uttering S to someone, intending for her listener to interpret S as S_1 but simultaneously committing herself—in whatever way our hypothetical speakers committed themselves above—to the truth of some other interpretation of S, namely S_2. The Gricean predicts that the strict truth and falsity of this utterance should correspond to S_1. I, in contrast, claim that it corresponds to S_2. If I'm right about this,

---

[23]While I have not run controlled tests on any of these judgments, extensive informal polling has proven remarkably consistent. Should the reader fail to feel the pull of the attested judgments, she is welcome to recast the argument as aimed at establishing a conditional claim: if these judgments prove robust, then here is what follows.

[24]See Saul (2012), chapters 1 & 2 for a helpful overview of these issues. Saul also considers a (rather different sort of) case involving demonstratives at pp. 111–14, but hesitates to draw any substantive conclusions from it.

[25]Cf. Michaelson (2016) and Stainton (2016) for related arguments.

the natural upshot is that we can say something without intending for our listener to come to believe, entertain, or otherwise recover that something.

So much for our initial case against the Gricean Theory of Reference. To be clear, this argument hardly rules out the possibility that intentions or other aspects of the speaker's mental state are what fixes literal reference, or other elements of literal meaning, in a context; all that it rules out is that it is specifically *Gricean* intentions which serve to do this. Before moving on to consider a non-Gricean alternative, however, let us first consider how the Gricean might try to account for these cases without giving up her theory.

4 Gricean Rejoinders

Response 1: The above judgments on (2) and (3) are mistaken. These utterances are, in fact, false—and thus constitute lies. An error theory can be offered to explain judgments like those attested to above.

Since the plausibility of this line of response will hinge on the details of whatever error theory is ultimately provided, it would be difficult, if not impossible, to definitively rebut this kind of response. Here, I will focus on what I take to be the most salient possibility.

This runs as follows: *qua* theorists, we have a tendency to excuse people, even to judge that they have not lied, when they intentionally say something false but can offer a plausible enough (though false) account of how what they said was true. In other words, the idea is that we systematically mistake a certain type of lie for a non-lie—even when we are privy to all the facts that should allow us to recognize the utterance as a lie. Various explanations might be offered for why we are prone to this sort of error, but one natural suggestion would be that, for whatever reason, we are apt to sympathetically project ourselves into the listener's perspective. So, if the listener is in no position to recognize that she is being lied to, and if she would remain in no position to so judge even on querying the speaker, we tend to take up her perspective and judge (incorrectly) that the speaker hasn't lied.

Consider how this sort of proposal will function in the two-Sams case: when called out for her mendacity, it is open to the speaker to claim that she really did intend for the listener to recover Sam_M as the referent. It simply slipped her mind that the listener would have no interest in having lunch with Sam_M! While we've stipulated that this was not the case, we can also

appreciate that the listener will be in no position to know this for certain; the speaker's claim is thus plausible enough. So, rather than correctly judging the speaker's utterance to be a lie, we are apt to allow the listener's perspective to unduly influence our judgments—leading us to judge, incorrectly, that this utterance was not only not a lie, but was perhaps even true.

But now consider what happens if we stipulate that the speaker unambiguously intends for her utterance of 'Sam' to refer to Sam_G in (3). In other words, suppose that the speaker flat-out intends to lie to the listener; no thought of how she could later construe this what she says as a non-lie ever crosses her mind as she formulates her utterance. In that case, it seems perfectly clear that the speaker has both said something false and that she lied.

The problem this poses for the error theory under consideration is that the speaker can offer an equally plausible story about how what she said was true in the latter case as she could in the former.[26] In other words, the listener's epistemic perspective remains unchanged across both versions of the case. Accordingly, this sort of error theory predicts that there should be no shift in our judgments on either lying or truth across these cases. Since our judgments do shift, we have reason to reject at least this particular error theoretic response to the challenge posed by cases of sneaky reference.

Response 2: The Gricean Theory, properly understood, actually predicts the desired referent in each case.

Implicit in the reasoning of the last section was the following thought: if one expects for one's listener to fail to identify O as the referent, and if one could easily have done something else to make it more likely that she would identify O as the referent, then one did not intend for her to recover O as the referent. But perhaps this sets the bar for intending too high; perhaps, as Neale (2004, p. 77) suggests, one can intend to ϕ so long as one doesn't take it to be *absolutely impossible* that one will succeed in ϕ-ing.

Applied to our cases of sneaky reference, this maneuver would allow the Gricean to predict that there is literal reference to the Camry and to Sam_M, respectively, in each of our cases. For in neither case should it strike the speaker as *impossible* for the listener to recover this object as the referent.

---

[26]In the event that the speaker needs to explain herself, suppose that the explanation occurs to her at precisely that the point when she is called on to do so.

Unfortunately, this strategy also threatens to undercut the Gricean's response to the Humpty Dumpty worry—for now her proposed constraints on referential intentions have become very weak. What's more, this strategy saddles the Gricean with two distinct referential intentions and two distinct literal referents—since granting that the speaker can intend to refer to the Camry or Sam_M by no means entails that she doesn't also intend to refer to the McLaren or Sam_G. So the Gricean owes us a story about what the notion of literal reference serves to explain if one can have two distinct literal referents.[27] This brings us to a third possible response.

Response 3: Cases of sneaky reference involve both an overt and a covert sort of speaker's meaning, corresponding to an overt and a covert sort of literal reference. Each plays a different explanatory role in an overall theory.

Applied to our cases of sneaky reference, the suggestion would be that, in (2), the speaker's utterance of 'that' overtly refers to the McLaren and covertly refers to the Camry. Likewise, in (3), the speaker's utterance of 'Sam' overtly refers to Sam_G and covertly refers to Sam_M. What's interesting about these cases is that they involve speakers who are simultaneously engaged in two different sorts of speech acts by means of a single utterance: an overt one of saying or asserting something to someone else and a covert one of talking to themselves. While perhaps a bit complicated, that observation hardly represents a fundamental challenge to the core Gricean thesis, **MEAN**.

It is worth noting straight off that this response relies on the disputed assumption that the Gricean can make sense of self-talk. For the sake of argument, I am happy to grant this.[28] Other worries remain, however.

First, the Gricean is effectively proposing that the relevant utterances are, in a substantial sense, non-audience-directed. I submit that this is simply the wrong way to describe (2) and (3). As best I can tell, there is no sense in which the speaker's utterance is self-directed in either of these cases. On

---

[27]See Buchanan (2018, p. 637) for a rather different criticism of Neale's proposal, along with a novel positive suggestion designed to accommodate cases where the speaker believes she will fail to get her meaning across: speaker's meaning depends not on the speaker's intentions *per se*, but rather on her intentional act of uttering something with a certain aim in mind. Unfortunately, this proposal will not help the Gricean to deal with cases of sneaky reference. The problem is that these are cases where the speaker's *aim* is to get her listener to identify an object which, intuitively, isn't the literal referent as the referent.

[28]See Grice (1969, pp. 174–7) for discussion.

the contrary, she is talking directly to the listener—just in a particularly tricky and opaque way. What she wants is, first, to get the listener to believe that p and, second, to go on the record, *publicly*, for the truth of some other proposition q. Neither of those acts looks to be self-directed.

Second, supposing that we accept the notion of 'covert' literal meaning, it is unclear what additional explanatory power we gain by positing 'overt' literal meaning as well. To see this, consider first the explanatory work that the so-called 'covert' notion of speaker's meaning is well-suited for: (i) it accurately predicts the strict truth conditions of utterances like (2) and (3), and (ii) it helps to explain why these utterances are not lies. Even if one were to dispute (i) on the basis of the judgments here being less clear than I have made them out to be, (ii) still looks like an important explanatory role for a notion of meaning to play.[29] Granted, nothing here entails that this is the only notion of meaning worth having around. But a notion of meaning that can help to explain such judgments looks to be one worth retaining.

Now let us consider what the 'overt' notion of meaning might help to explain. If the truth conditions of (2) and (3) really are unclear, then positing this sort of meaning would help to explain why. But we can just as easily explain this by starting with 'covert' meaning and then noting that no reasonable listener is likely to recover this meaning in the relevant contexts. Being attuned to this, some will tend to adjust their confidence in their truth/falsity-judgments downwards. Given the availability of such a pragmatic explanation, this looks like a rather thin case for positing two distinct types of literal meaning.

Alternatively, one might claim that overt meaning is necessary to explain communicative potential. The problem, however, is that it's highly dubitable that anything like successful communication is possible in cases of sneaky reference. In ordinary communicative cases, on the other hand, the speaker's

---

[29]Cf. Saul (2012), Michaelson (2016), Stainton (2016), and Borg (2019). The argument becomes more complicated if we allow that utterances can express 'clouds' of propositions, which then might then figure into explanations of lying. Note, however, that an appeal to clouds will hardly count as a defense of the Gricean Theory should propositions that fail to conform to $\mathbf{MEAN}_L$ be allowed into the cloud, as would seem to be the natural suggestion when it comes to cases of sneaky reference. What's more, the two extant cloud-tolerant accounts of lying—those of Saul (2012) and Stokke (2016, 2018)—both suffer from serious drawbacks unrelated to anything presently under discussion. See Stokke (2016, 2018) for criticism of Saul's proposal and van Elswyk (2020) for criticism of Stokke's.

overt and covert meanings will coincide—so we are free to run our explanation of communicative potential in these cases purely in terms of 'covert' meaning.

5   The Direct Expression View of Reference

Having cast doubt on the Gricean Theory, it remains for me to offer a viable alternative. Like the Gricean, I suggest that we think of literal reference as determined by the speaker's mental state. In contrast to the Gricean, however, I don't think the relevant aspect of the speaker's mental state is audience-directed. Instead, I would suggest that literal meaning is the result of the speaker's intention to *express a thought*, where this is to be understood as vocalizing that thought by means of some appropriate words. Literal reference, in turn, depends on intentions to vocalize the object-directed sub-parts of one's thoughts.[30] Given these basic commitments, I will call my alternative the 'Direct Expression View'. We'll return to the expression part of this shortly. First though, I want to flesh out what it means to use some 'appropriate words' to express a thought.

I will take as my starting point Strawson (1950)'s suggestion that 'To give the meaning of an expression...is to give general directions for its use to refer to or mention particular objects or persons' (p. 327). Put slightly differently, we might think of referring terms as being associated, via convention, with rules specifying the sorts of things to which they can be used to literally refer in a given context. So, for instance, the name 'Sam' can be used to literally refer to an individual named 'Sam'.[31] One might *try* to use the term to refer to someone who is named something else—say, 'Jess'. Indeed, one might even succeed, in the right circumstances, in communicating something about Jess by means of an utterance of the name 'Sam'. But one cannot *literally* refer to someone who isn't named 'Sam' with an utterance of the name 'Sam'.

Let us call these constraints on successful literal reference the 'presuppositions' associated with a given expression. When one utters a referential term, a candidate-referent will either *satisfy* or *fail to satisfy* the relevant presuppositions. Where a candidate-referent fails to satisfy those presuppositions, literal reference fails. For ease of exposition, I will assume that such

---

[30]Alternatively, if we can have attitudes towards bare objects, then one might be expressing the object of such an attitude. See Grzankowksi (2016) for recent discussion.

[31]Cf. Schoubye (2017). See also Gray (2015, 2017) for discussion of how complex this seemingly-simple property may turn out to be.

reference failure makes utterances of simple, subject-predicate sentences involving referential terms *false*.

What remains is to specify what makes a particular object the candidate-referent for some particular utterance of a referential term. Here, Strawson (1950) is less helpful; while he speaks of referential expressions as being 'used' to refer to objects, he never elaborates on what makes a particular utterance of a term counts as one sort of use as opposed to another.[32] As a first step, I suggest taking a page from the Kaplanian playbook: it is the fact that the speaker intends to use e to refer to O that makes it the case that e is being used to refer to O. Or, more accurately, it is this fact which makes O the candidate-referent; the utterance of e could still fail to literally refer to O in the event that O fails to satisfy the relevant presuppositions.

Now we need a non-Gricean way of understanding referential intentions. This is where the notion of *expressing* comes in. My suggestion is that an intention to refer to O by an utterance of e should be understood as: an intention to use e to express an O-directed sub-part of an O-involving thought.[33] What sort of act-type is this? As a start, we might say that it is whatever is common to meaningful utterances of referential terms when talking to particular people, to ourselves, to no one in particular, or to no one at all. What is common to all of these, I would suggest, is that each is an instance of *voicing a thought* or *putting a thought into words*.[34] When a thought is object-directed, then voicing that thought necessarily involves a subsidiary act of voicing the object-directed sub-part of that thought.

---

[32]Later Strawson, e.g. Strawson (1964), might well have appealed to a modified version of Gricean speaker's meaning—at least for communicative speech acts.

[33]It is also worth noting that the theory is neutral with regard to most aspects of the nature of the thoughts involved. For example, both internalists and externalists about the contents of thoughts should be free to endorse the Direct Expression Theory.

[34]See, *inter alia*, Bach & Harnish (1979), Davis (2003), Green (2007), Bar-On (2013), and Radulescu (2019) for alternative understandings of *expression*. Some salient points of contrast: Bach & Harnish (1979, p. 39) famously analyze expression in explicitly Gricean terms, thereby requiring both audience-directedness and reflexivity. Davis (2003), on the other hand, explicitly rejects many aspects of the Gricean framework. Nonetheless, Davis preserves a kind of audience-directedness in his understanding of expression—as does Green, who follows Davis in this respect—via his analysis of expression as involving the intentional production of an 'undisguised indication' of a belief (Davis 2003, p. 49). Since intentional indications are plausibly never disguised to the agent producing them, this way of understanding expression is still fundamentally audience-directed.

Still, we might hope for greater clarity than this. So let us try a different tack. Consider once more Strawson's idea that linguistic meanings are something like rules for using words and sentences in context. How exactly should we understand such rules? One possibility would be to think of each sentence as being associated with a set of conditions in which it can be appropriately uttered; words, in turn, would systematically contribute to such appropriateness conditions. While pleasingly simple, this way of thinking about things risks obscuring an important fact: plausibly, and particularly when ambiguous or context-sensitive language is involved, the speaker must commit herself to the world's being some more specific way than what is reflected in such broad appropriateness conditions. In other words, she is expected to go on the record for the world's being some specific way—regardless of whether that record is anything more than in-principle accessible to her interlocutor, and regardless of whether the speaker intended for it to be such.[35]

To better reflect this requirement, I suggest thinking of relevant rules as something like exhaustive sets of permissions: by uttering U in a context C, one is free to commit to any acceptable interpretation of U in C. That is, one is free to commit to the world's being as any of p_1 or p_2...or p_n would have it. But one must commit to one of these.[36] Likewise, by uttering a referential term e in C, one is permitted to literally refer to one of a specified range of objects, $O_1...O_n$. For example, in uttering the name 'Sam', one is free to literally refer to any Sam that one likes. But again, one must commit to one of these.[37] Otherwise, one is merely mouthing words.

---

[35]Note that, if King (2018) is right and certain referential terms can be used felicitously even without full specification of a referent, then we'll need to understand this notion of 'going on the record for the world's being some specific way' in a way compatible with this potential looseness—that is, as requiring specificity, but not necessarily *full* specificity. For ease of exposition, I will suppress this complication in what follows.

[36]I am ignoring the possibility that we might sometimes use a single utterance to say different things to different people. For discussion, see Egan (2009).

[37]We should also insist on these sentence-level permissions being built up, systematically, from expression-level permissions and modes of combination. A compositional theory along these lines in indeed possible, though best pursued elsewhere. See Michaelson & Woods (2020).

It is also worth noting that the permissions associated with particular expressions may turn out to be more considerably more complex than any we have considered here. See von Fintel & Gillies (2011) for a view about the meaning of 'might' that can be productively understood along these lines.

Typically, of course, we say things in order to try to get our interlocutors to believe, or at least to entertain, what we've said—not just to go on the record for the world's being such-and-such a way. On the present picture, what this means is that we're usually going to do our best to make our interpretive commitments evident to the listener. The mistake the Gricean makes, I've argued, is to treat this feature of cooperative speech acts as an essential feature of meaningful speech simpliciter. In contrast, I have suggested that what is common between cooperative and non-cooperative utterances of declarative sentences is that both involve committing oneself in a particularly strong manner to the world's being a certain way—such that one becomes lie-prone with respect to that content. One undertakes such a commitment in part via subsidiary commitments to particular interpretations of one's words. If one is confused about the rules for using those words, or if one tries to use them in some deviant manner, then those words will lack a literal meaning; one will have tried, but failed, to voice a thought.[38]

Putting the pieces together, we can now clarify the proposed view of expression:

> **EXPRESS**   To *express the thought that* p by an utterance U is to intend to commit oneself to the world's being p in uttering U, where p is an acceptable interpretation of U in the relevant context.

One can thus fail to express the thought that p if one intends to commit oneself to the world's being p in uttering U, but then one chooses some U that doesn't have p as an acceptable interpretation in the relevant context. So when I say that a speaker *intends to express the thought that* p with a given utterance, what I mean is that she both intended to commit herself to the truth of p in uttering U and that she intended to select a U for which p is

---

[38]See, *inter alia*, Brandom (1994) for an alternative commitment-based account of meaning and reference. Saliently, for Brandom, an 'assertional commitment' requires that the speaker at least intend to make this commitment transparent (Ibid., pp. 167–8). This, I take it, is Brandom's analogue of **MEAN**. My view imposes no such requirement. For a much more similar account to my own, see Cumming (2015); in contrast to Brandom, Cumming allows that reference can be fixed by the speaker's 'private commitments'.

an acceptable interpretation in context.[39] Intentions to refer, in turn, are to be understood as sub-parts of intentions to express object-directed thoughts.

One final bit of the picture is still lacking, however: to explain our cases of sneaky reference, we must more fully specify the presuppositions associated with names and demonstratives. We've already seen how to treat names like 'N': an utterance of 'N' presupposes that the referent bears the name 'N'. An utterance of the demonstrative 'this' I take to presuppose that the referent is both relatively proximal and in the region of space indicated by the accompanying ostensive gesture, should there be one.[40] 'That' functions analogously, but presupposes that the referent is relatively distal rather than proximal. It is worth emphasizing that, in the absence of an accompanying gesture, both 'this' and 'that' will turn out to be *extremely* flexible devices of reference; when paired with a gesture, each becomes much less flexible.

Now let us return to our cases of sneaky reference. In the Camry-McLaren case, the speaker intended for her utterance of 'that' to refer to the Camry. That is, she intended to express a Camry-directed thought. Since the Camry was both distal and indicated by the speaker's gesture—i.e. it was among the objects in the region of space indicated by the gesture—it satisfied the relevant presupposition and she succeeded in expressing this thought. So the speaker's utterance of 'that' is predicted to literally refer to the Camry.

In the second case, the two-Sams case, the speaker intended for her utterance of 'Sam' to refer to Sam_M. That is, she intended to express a Sam_M-directed thought. Since Sam_M bears the name 'Sam', the relevant presupposition is satisfied and the speaker succeeded in expressing that thought. Thus, her utterance of 'Sam' is predicted to literally refer to Sam_M.

The Direct Expression View thus successfully predicts the intuitive literal referent in each of our cases of sneaky reference. But at what cost? Does it, for instance, turn out to be a Humpty Dumpty theory?

It does not. Consider once more Reimer's Fido-Spot case. The worry here was that, on the Kaplanian theory, we looked stuck predicting that the

---

[39]More deserves to be said about what it means to commit oneself to the truth of something, but I will have to leave that for future investigation. Very briefly, I take this to be similar to the sense in which we commit ourselves to the truth of p when we come to believe p—the salient difference being that belief is not (or not typically) volitional.

[40]I take this notion of 'indication' from Lascarides & Stone (2009). Basically, the idea is that gestures are conventionalized ways of picking out regions of space, along with all of the various objects and properties within them. See also Elbourne (2008) and Sherman (2015) for discussion of the relevant notions of distility and proximity.

speaker's utterance of 'this' literally referred to Fido even when Fido had run away and the speaker had accidentally gestured at Spot. The Direct Expression View, in contrast, makes no such prediction. Fido fails to satisfy the relevant presuppositions: having run away, Fido is neither to be found among the objects or properties indicated by the speaker's ostensive gesture, nor is he particularly proximal. So, on the present view, the speaker tries to use 'this' to literally refer to Fido. But she fails to actually do so.

There is a further question though: does the speaker's utterance of 'this' literally refer to Spot? Reimer (1991, p. 182) claims that it does, and I am inclined to agree. I would note, however, that I am less certain of this than I am that the speaker's use of 'this' does not refer to Fido. Regardless, this would seem to constitute a problem for the Direct Expression View as stated—for the fact that the speaker's utterance of 'this' fails to refer to Fido in no way entails that it literally refers to Spot.

A relatively easy fix is in the offing, however. Let us follow Bach in positing that a speaker in a case like Reimer's will have at least two relevant intentions: (i) an intention to refer to Fido, and (ii) an intention to refer to whatever dog she ostends with her embrace. Now let us further posit that, in cases where multiple referential intentions are in play, the presuppositions associated with a particular referential expression serve to *filter* these intentions. So, where a speaker intends to refer to both O_1 and O_2—not realizing that these are distinct—but where only O_1 satisfies the relevant presuppositions for e, then the utterance of e literally refers to O_1. If neither satisfies the relevant presuppositions, or if both do, then reference fails.[41]

Here is how this slightly modified version of the theory can handle the Fido-Spot case: the speaker intends both to refer to Fido and to Spot. But since she ostends only Spot, only Spot satisfies the presupposition associated with her utterance of 'this' in conjunction with her gesture. So 'this' literally refers to Spot, and only to Spot—thus vindicating Reimer's judgment here.

Let us take stock. Like the Gricean Theory of Reference before it, the Direct Expression View looks to be genuinely non-Humpty Dumptian. In contrast to the Gricean Theory, however, the Direct Expression View is also capable of explaining literal reference in our cases of sneaky reference. This would seem to constitute a significant advantage for the Direct Expression View. As with any new theory, however, worries are bound to arise.

---

[41]I explore this notion of a *filter* in greater detail, and offer a formal characterization of that notion, in Michaelson (2020).

6 Objections and Replies

Objection 1: The Direct Expression View makes non-cooperative meaning just as basic as cooperative meaning. But that's implausible. Rather, non-cooperative meanings like those evinced in cases of sneaky reference are metaphysically dependent on cooperative meanings.

Reply   The Direct Expression View allows certain non-cooperative uses of language, such as cases of sneaky reference, to count as *meaningful* in the sense of being evaluable for truth or falsity. That, however, by no means entails that non-cooperative uses are on par with cooperative uses in terms of what grounds the standing meanings of expressions. Following Lewis (1969, 1975), we might posit that, for a language to coalesce at all, something like a convention of truthfulness and trust must obtain among a sufficient number of speakers. This would seem to grant a kind of metaphysical priority to cooperative uses. But that priority is perfectly compatible with there being tricksters willing to take advantage of these conventions once they have been established—as Lewis was himself was perfectly willing to grant.

Alternatively, the objector might have had a different sense of 'dependence' in mind: perhaps cooperative uses serve to establish what a particular expression is *for*, to establish its teleology, in a way that uncooperative uses do not. Again, this is perfectly compatible with the Direct Expression View. That theory is meant to explain how and when we can use expressions to mean specific things *once their rules of use have already been established*. Perhaps cooperative uses really do establish what an expression is for in a way that uncooperative uses do not. All that I have been concerned to show is that sufficiently complex speakers can learn to exploit the rules of languages like our own so as to say things which are strictly speaking true, yet which are explicitly designed to be misunderstood by their listeners.

Objection 2: The Direct Expression View divorces semantic or assertoric content from either actual or expected communication. That's bad. The theoretical role of such content is to help to explain communication.

Reply   This objection presupposes a certain picture of the explanatory role of semantic, assertoric, or truth-conditional content that I am inclined to reject. Recent advocates of this sort of picture have suggested that such content must either be available to (i) the actual (Stokke 2010) or (ii) to a

suitably-idealized (Neale 2004, King 2014) listener. This content can then be said to 'explain' communication in the following sense: since that content either was or ought to have been available to the listener, there is no mystery as to how the listener could be expected to recover the relevant content.

It seems to me, however, that there is no real explanation on offer here. To better understand communication, we might hope to better understand the mechanisms by which listeners do, and ought to, interpret utterances. This sort of proposal effectively stipulates a correlation between the outputs of our (potentially idealized) interpretive mechanisms and a particular sort of content—but that is hardly to cast light on the nature of the mechanisms themselves. The following strikes me as a better start: listeners typically make use of the presuppositions associated with words and utterances, in addition to their knowledge of the background situation, to try to figure out what the speaker has committed herself to as the literal meaning of her utterance. Then, they engage in further reasoning about the utterance. All of that, of course, is perfectly in keeping with the Direct Expression View.

What's more, it is highly dubitable that semantic or assertoric content must be recoverable by either the actual or a suitably-idealized listener.[42] We saw above that there is a whole class of speech acts which have perfectly good literal meanings (or semantic or assertoric contents, if one prefers) that even a suitably-idealized listener will be in no position to recover. These acts involved speakers who were willing to go to some lengths to minimize the chance of their listeners recovering the literal meanings of their utterances. This is why I think we would do well to think of such speech acts as fundamentally *non-cooperative*; they are simply not aimed at communication.

Now, one could grant all this but propose that the terms 'semantic' or 'assertoric' content be reserved for some other notion than what we have been calling 'literal meaning'. Perhaps such content only attaches to cooperative speech acts. I have no real bone to pick with this suggestion, other than to point out that this notion of content seems likely to be explanatorily inert. The notion of literal meaning that I have been developing looks well-suited to help explain (i) our truth-value judgments across both cooperative and non-cooperative utterances, (ii) our judgments on lying across both these sorts of utterances, and (iii) communication in cooperative cases, by setting a target for the listener to recover.[43] It is unclear to me what this other no-

---

[42]For additional, independent arguments to this effect, see Nowak & Michaelson (2020).

[43]For a defense of this basic picture of communication, see Pagin (2008).

tion of content serves to add to our explanations of (i)–(iii)—or what other phenomena beyond these it might help to explain. So, at present at least, we seem to lack anything like a compelling reason to posit such a notion.

Objection 3: The Direct Expression View offers only a superficial treatment of names and demonstratives. For example, no account of non-rigid uses of these terms has been offered.

The objector here is certainly correct to point out that the Direct Expression View, as I have sketched it above, offers nothing like a full semantic treatment of names or demonstratives. She is wrong, however, to view this as an objection. The view outlined above is a metasemantic view, not a semantic one; it is a view about what determines meaning in context, not about what particular sorts of terms mean in context. In fact, the view has been explicitly designed to play nicely with a wide range of semantic theories for names and demonstratives—even with those from outside the direct reference tradition. This may not seem obvious, however, since I talked freely above about names and demonstratives having 'referents'. While that is a natural way of thinking about things, it isn't strictly essential to the view. To see why, it will help to work through some examples.

Consider first a theory like Elbourne (2008)'s, on which demonstratives are taken to denote complex properties that Elbourne calls 'individual concepts'. Prescinding from a number of interesting and important technical details, Elbourne derives these properties by assuming that any token demonstrative comes paired with an index i and relation R. To account for the appearance of reference and rigidity, Elbourne claims that:

> [I]t will frequently happen...that the property derived from i and R will be the property of being identical to a particular person or thing, so that the same person or thing will be the value of the demonstrative (when it has a value) whatever evaluation situation is considered. (Ibid., p. 430)

Non-rigid interpretations of demonstratives, in contrast, are generated by allowing R to be instantiated by something other than the identity relation.

The important thing for our purposes is that Elbourne never offers a theory of what determines the value of the index i in context. That is a task for a metasemantic theory, and hence one to which the Direct Expression

View is well-suited: the value of i is determined in both rigid and non-rigid contexts by the thought that the speaker is trying to express. In fact, the view can be used to help determine the value of R as well: R will be the identity relation when the speaker is expressing a rigid thought, and something else when she is expressing a non-rigid one.

Analogous considerations will hold in the case of names. Take, for instance, Graff Fara (2015)'s predicativist view of names, according to which bare argument-place uses of names are semantically equivalent to incomplete definite descriptions (Ibid., p. 97). This equivalence allows Graff Fara to explain the apparent rigidity of such uses, since indefinite descriptions are typically rigid. But this explanation comes at a cost: like other Russellians, Graff Fara needs some way of completing the description in context, such that it can serve to isolate a single object.[44] Graff Fara never offers a theory of what provides such completions. Here again, the Direct Expression View looks poised to help. Taking a cue from Elbourne, we might allow the completing property to be *being identical to the value of some index i*, with the value of i being identical to whatever object the speaker is expressing a thought about—supposing that this object is in fact called '*N*'.

In summary: the objector was right to point out that the Direct Expression View provides neither a semantics for names nor one for demonstratives. But at this point in the development of semantics, it should probably be considered a theoretical virtue for a given metasemantic view to be compatible with a wide range of semantic theories.

7 Conclusion

Above, I pointed to a class of cases—involving agents who are deliberately speaking so as to be misunderstood—which poses a serious challenge for the core Gricean thesis about meaning, a thesis which has been adopted by an extraordinarily wide range of contemporary theorists. The basic problem is that these utterances are fundamentally non-cooperative, whereas the core Gricean thesis has it that, for speech to be meaningful, it must be cooperative in some basic sense. Since so much of our speech is cooperative in the Gricean sense, this idealization proves harmless in most contexts. But it threatens to

---

[44]Graff Fara notes that she is inclined to reject the claim that any unique completion need be singled out by the context (Ibid., p. 98). Still, even if there is a class of acceptable completions we need some way of determining the bounds of this class.

significantly distort our understanding how language works if we bake it into our theory of meaning at the outset. I have argued that, while there is no easy fix once we have adopted this idealization, the idealization can in fact be dropped without any catastrophic consequences.

The arguments above focused on literal meaning, and even more specifically on literal reference. But the point carries over to speaker's meaning and speaker's reference as well—at least on the assumption that we have to speaker mean what we say. For, if one can say p in spite of intending for one's listener to recover q, then one must be able to speaker mean p in spite of this as well. If we could not, then sneaky reference and other instances of speaking so as to be misunderstood would be impossible. Since these are possible, then either we must reject this claim about speaker's meaning or else we must reject the claim that one must mean what one says.

I would suggest going the former route, though doing so leaves us in need of an alternative account of speaker's meaning. Here is a suggestion: the rules of a language can be viewed as giving rise to a peculiar sort of game, one in which your moves are only ever partially visible to other players. Imagine a game like Risk, but where players put their pieces under opaque cups before placing them on the board. When you occupy a territory, that much is clear; your cup is now there. What's unclear is just how many pieces you've placed—though, of course, some constraints on this number will be known to all. In such a game, there *is* a fact of the matter about how many pieces you have where. Sometimes, when you are playing a cooperative version of the game, you may go to some lengths to try and make clear to the other players how many pieces you've placed where. Other times, you might not make any such effort. And sometimes you might even try to deceive your fellow players about how exactly your pieces are arranged.[45]

The Direct Expression View treats speaker's meaning roughly on this model: such meaning involves committing oneself to a particular interpretation of one's utterance, or going on the record in a particular way, even when no one else is in a position to recognize this commitment. An open question at this point is whether this sort of speaker's meaning will suffice for us to explain hinting and implicating, or whether something more like Gricean speaker's meaning is required to explain these. I won't try to settle the matter here, but I will note the following: if my arguments above prove correct, and yet the core Gricean thesis really does capture something essen-

---

[45]The idea for games of this sort comes from Iain Banks' novel *The Player of Games.*

tial to hinting and implicating, then speaker's meaning may well turn out to be a far more disunified notion than it is standardly taken to be.[46]

References

Avramides, Anita. 1989. *Meaning and mind: An examination of a Gricean account of meaning.* Cambridge, Massachusetts: MIT Press.

Bach, Kent. 1992. Intentions and demonstrations. *Analysis* 52. 140–146.

Bach, Kent. 2005. Context *ex Machina.* In Zoltán Szabó (ed.), *Semantics versus pragmatics*, 15–44. Oxford: Oxford University Press.

Bach, Kent & Robert M. Harnish. 1979. *Linguistic communication and speech acts.* Cambridge: MIT Press.

Bar-On, Dorit. 2013. Expressive communication and continuity skepticism. *Journal of Philosophy* 110(6). 293–330.

Bennett, Jonathan. 1976. *Linguistic behaviour.* Cambridge: Cambridge University Press.

Borg, Emma. 2019. Explanatory roles for minimal content. Forthcoming in *Noûs.*

Brandom, Robert B. 1994. *Making it explicit: Reasoning, representing, and discursive commitment.* Cambridge, Massachusetts: Harvard University Press.

Buchanan, Ray. 2014. Reference, understanding, and communication. *Australasian Journal of Philosophy* 92(1). 55–70.

Buchanan, Ray. 2018. Intention and the basis of meaning. *Ergo* 5(23). 629–644.

Cumming, Samuel. 2015. The dilemma of indefinites. In Andrea Bianchi (ed.), *On reference*, 335–349. Oxford: Oxford University Press.

Davidson, Donald. 1986. A nice derangement of epitaphs. In R. Grandy & R. Warner (eds.), *The philosophical grounds of rationality*, chap. 157-174. Oxford University Press.

Davis, Wayne A. 2003. *Meaning, expression and thought.* Cambridge: Cambridge University Press.

Egan, Andy. 2009. Billboards, bombs, and shotgun wedddings. *Synthese* 166. 251–279.

Elbourne, Paul. 2008. Demonstratives as individual concepts. *Linguistics and Philosophy* 31(4). 409–466.

van Elswyk, Peter. 2020. Deceiving without answering. Forthcoming in *Philosophical Studies.*

von Fintel, Kai & Anthony S. Gillies. 2011. 'might' made right. In Andy Egan & Brian Weatherson (eds.), *Epistemic modality*, 108–130. Oxford: Oxford University Press.

Glüer, Kathrin & Peter Pagin. 2003. Meaning theory and autistic speakers. *Mind and Language* 1(18). 23–51.

Graff Fara, Delia. 2015. Names are predicates. *Philosophical Review* 124(1). 59–117.

Gray, Aidan. 2015. Lexical individuation and predicativism about names. *Thought* 4(2). 113–123.

Gray, Aidan. 2017. Names in strange places. *Linguistics and Philosophy* 40(5). 429–472.

Green, Mitchell S. 2007. *Self-expression.* Oxford: Oxford University Press.

Grice, H. Paul. 1957. Meaning. *Philosophical Review* 66(3). 377–388.

Grice, H. Paul. 1968. Utterer's meaning, sentence-meaning, and word-meaning. *Foundations of Language* 4. 225–242.

Grice, H. Paul. 1969. Utterer's meaning and intentions. *Philosophical Review* 78(2). 147–177.

Grice, H. Paul. 1989. Logic and conversation. In *Studies in the way of words*, 22–40. Cambridge: Harvard University Press.

Grzankowksi, Alex. 2016. Attitudes towards objects. *Noûs* 50(2). 314–328.

Kaplan, David. 1989. Afterthoughts. In Joseph Almog, John Perry & Howard Wettstein (eds.), *Themes from Kaplan*, 565–614. Oxford: Oxford University Press.

Kaplan, David. 1990. Words. *Proceedings of the Aristotelian Society* 64. 93–119.

King, Jeffrey C. 2014. Speaker intentions in context. *Noûs* 48(2). 219–237.

King, Jeffrey C. 2018. Strong contextual felicity and felicitous underspecification. *Philosophy and Phenomenological Research* 97(3). 631–657.

Kotatko, Petr. 1998. Two notions of utterance meaning. *Proceedings of the Aristotelian Society* 98. 225–239.

Kripke, Saul. 1977. Speaker's reference and semantic reference. In Peter French, Thomas Uehling & Howard Wettstein (eds.), *Midwest studies in philosophy, vol. 2*, 255–276. Minneapolis: University of Minnesota Press.

Lascarides, Alex & Matthew Stone. 2009. A formal semantics of gesture. *Journal of Semantics* 393–449.

Laurence, Stephen. 1998. Convention-based semantics and the development of language. In P. Carruthers & J. Boucher (eds.), *Language and thought: Interdisciplinary themes*, 201–217. Cambridge University Press.

Lewis, David. 1969. *Convention*. Cambridge: Harvard University Press.

Lewis, David K. 1975. Languages and language. In Keith Gunderson (ed.), *Language, mind, and knowledge. Minnesota studies in the philosophy of science, vol. 7*, 3–35. Minneapolis: University of Minnesota Press.

Lewis, David K. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8(3). 339–359.

MacKay, Alfred F. 1968. Mr. Donnellan and Humpty Dumpty on referring. *Philosophical Review* 77(2). 197–202.

McGinn, Colin. 1981. The mechanism of reference. *Synthese* 49(2). 157–186.

Michaelson, Eliot. 2016. The lying test. *Mind and Language* 31(4). 470–499.

Michaelson, Eliot. 2020. Towards a grand unified theory of reference. Manuscript.

Michaelson, Eliot & Jack Woods. 2020. Building character. Manuscript.

Neale, Stephen. 1992. Paul Grice and the philosophy of language. *Linguistics and Philosophy* 15. 509–559.

Neale, Stephen. 2004. This, that and the other. In Marga Reimer & Anne Bezuidenhout (eds.), *Descriptions and beyond*, 68–182. Oxford: Oxford University Press.

Nowak, Ethan & Eliot Michaelson. 2020. Who's your ideal listener? Forthcoming in *Australasian Journal of Philosophy*.

Pagin, Peter. 2008. What is communicative success? *Canadian Journal of Philosophy* 38. 85–116.

Radulescu, Alexandru. 2019. A defence of intentionalism about demonstratives. Forthcoming in *Australasian Journal of Philosophy*.

Reimer, Marga. 1991. Do demonstratives have semantic significance? *Analysis* 51(4). 177–183.

Reimer, Marga. 1992. Three views of demonstrative reference. *Synthese* 93(3). 373–402.

Saul, Jennifer. 2012. *Lying, misleading, and what is said: An exploration in philosophy of language and in ethics*. Oxford: Oxford University Press.

Schiffer, Stephen. 1972. *Meaning.* Oxford: Oxford University Press.

Schiffer, Stephen. 1981. Indexicals and the theory of reference. *Synthese* 49. 43–100.

Schoubye, Anders J. 2017. Type-ambiguous names. *Mind* 126(503). 715–767.

Searle, John R. 1969. *Speech acts.* Cambridge: Cambridge University Press.

Sherman, Brett. 2015. Constructing contexts. *Ergo* 2(23). 581–605. Manuscript, Brandeis.

Stainton, Robert J. 2016. Full-on stating. *Mind and Language* 31(4). 395–413.

Stine, Gail C. 1978. Meaning other than what we say and referring. *Philosophical Studies* 33(4). 319–337.

Stokke, Andreas. 2010. Intention-sensitive semantics. *Synthese* 175(3). 383–404.

Stokke, Andreas. 2016. Lying and misleading in discourse. *Philosophical Review* 125(1). 83–134.

Stokke, Andreas. 2018. *Lying and insincerity.* Oxford: Oxford University Press.

Strawson, Peter F. 1950. On referring. *Mind* 59(235). 320–344.

Strawson, Peter F. 1964. Intenion and convention in speech acts. *Philosophical Review* 73(4). 439–460.

Vlach, Frank. 1981. Speaker's meaning. *Linguistics and Philosophy* 4. 359–391.

Weatherson, Brian. 2002. Misleading indexicals. *Analysis* 62(4). 308–310.