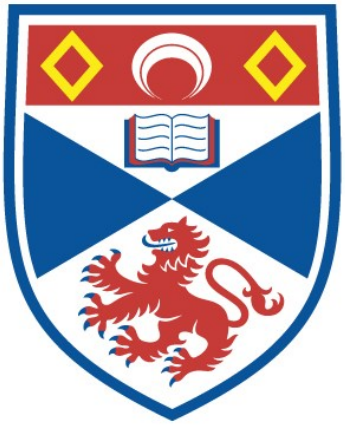
What people say: publicity without semanticity

Victor Tamburini

A thesis submitted for the degree of PhD
at the
University of St Andrews



2024

Candidate's declaration

I, Victor Tamburini, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 74,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in September 2020.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date: 19/04/2024 Signature of candidate:

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

Date: 19/04/2024 Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Victor Tamburini, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Date: 19/04/2024 Signature of candidate:

Date: 19/04/2024 Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Victor Tamburini, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date: 19/04/2024 Signature of candidate:

General acknowledgements

I wish to thank my supervisors Derek Ball and Jessica Brown for their unwavering support throughout my PhD. Their philosophical insights permeate every chapter of this dissertation. I am also grateful to Manuel García-Carpintero and François Recanati for helpful discussions during my visits in Barcelona and Paris. Thanks are due to the participants of the Arché Language & Mind seminar for excellent comments on my work, in particular Justin D'Ambrosio, Daniel Garibay-Garcia, Jean Gové, Lixiao Lin, Rasmus Overmark, and Simon Prosser. Finally, I am eternally grateful to Xintong Wei for her support and friendship throughout my PhD.

Funding

This work was supported by the University of St Andrews (Department of Philosophy).

Thesis abstract

When a speaker utters a declarative sentence, we may report her speech indirectly: 'She said that *p*'. What determines the content reported as what is said (*p* in 'she said that *p*')? In this dissertation I argue in a novel way against *semanticity*, the view that what speakers say is linguistically determined. I focus first on demonstrative pronouns ('this', 'that'), showing that attempts to construe their contributions to what is said as determined by a linguistic rule fail. Next I present properties of what is said which seem to support *semanticity*, notably the publicity of what is said. Publicity is the property of being independent from both the speaker's intentions and the audience's actual interpretation. I argue that the publicity of what is said is not best explained by *semanticity*. I go on to present a non-semantic explanation of the publicity of what is said, which leads me to a general view of the determination of the content of speech acts.

Table of contents

Introduction

When a speaker utters a declarative sentence, we may report her speech indirectly: 'She said *that p*'. With indirect speech we focus on the information conveyed by the uttered sentence - *that p* - rather than the sentence itself. Some philosophers have pointed out that indirect 'say that' reports are very loose, in the sense that various pieces of information more or less explicitly conveyed by the uttered sentence can figure in the subordinate that-clause (Cappelen & Lepore 1997). Nevertheless, there is an ordinary sense of 'say' according to which what a speaker says is closely related to the properties of the uttered sentence, in contrast with what the speaker merely suggests (hints, insinuates) for example. With a given sentence, anything can be suggested or insinuated depending on the context of utterance; but not everything can be *said*. What is said stands out among the contents of an uttered sentence, by being constrained by the syntax and meaning of the sentence.

The idea that what is said is constrained by the syntax and meaning of the uttered sentence is consistent with a wide range of theoretical views. But according to the strongest interpretation of 'constrained', what is said is *fully determined* by syntax and meaning. If we follow contemporary semantics in thinking that linguistic meaning is compositional, this interpretation yields the thesis that what is said is fully determined by the meaning of the words contained in the sentence and compositional rules (rules that churn out the meaning of a complex expression, given its syntactic structure and the meaning of the simple expressions it contains). This thesis may be called the semanticity of what is said - or *semanticity* for short.

Why would one choose the strongest interpretation of 'constrained' and endorse *semanticity*? There are several factors to take into account here. First, from the point of view of someone aiming to build a theory of linguistic meaning - a semantic theory - the more the theory explains the better. If semantic theory can go all the way to predicting the main content of the utterance of a declarative sentence - what the speaker says - that it can do so is a virtue. Secondly, some of the basic data of contemporary semantic theory are the judgements of ordinary speakers on the truth or falsity of what is said with sentences. Hence, what is said is the object of the pre-theoretical judgements that semantic theory is built upon. If on the one hand *what is said* is used as evidence to build a theory of linguistic meaning, but on the other hand the object that the theory is built upon is not entirely determined by linguistic meaning, difficult methodological questions loom for this theory. Thirdly, mainstream semantic theories are built on the idea that giving the meaning of a declarative sentence is giving its *truth-conditions* - the conditions under which the sentence is true or false (Davidson 1967, Heim & Kratzer 1998, Larson & Segal 1995, Lewis 1970). Since what a speaker says with a declarative sentence can be true or false, it can play the role of meaning *qua* truth-conditions. The most straightforward view to take here is that the meaning-giving truth-condition is just what the speaker says. Of course, it is possible to hold that the meaning of an uttered sentence

is a truth-condition, yet one that is distinct from what the speaker says. But this more complicated view is fraught with difficulties (Borg 2004, 2019).

Semanticity has been refined to deal with context-sensitive expressions such as 'I' and 'Today'. Context-sensitive expressions are expressions whose meaning dictates that the contribution to what is said made by their utterance varies with the context of utterance. The problem for *semanticity* posed by context-sensitive expressions is that their contribution to what is said solicits more than their meaning. Take the word 'I': it is used to say things about different people in different contexts. Of course, the feature of context that matters is the identity of the speaker. And this is just what one knows when one knows the meaning of 'I': that 'I' refers to whoever utters it. The meaning of 'I' thus tells us how what is said with an utterance of 'I' varies with the context of utterance. What is said with an utterance of 'I' - i.e. the person referred to - is then not determined by the meaning of 'I' alone: it is determined by the meaning of 'I' *plus* the aspect of context invoked by this meaning - i.e. the identity of the speaker. In order to retain full generality, *semanticity* must then be amended: what is said with the utterance of a sentence is fully determined by (i) the meaning of the words contained in the sentence; (ii) rules of meaning-composition; and *(iii) the aspects of context of utterance invoked by the meaning of the words.*

This amended version of *semanticity* is in turn challenged by sentences with the following profile: some aspects of what is said with them are a matter of context, but the relevant contextual determinants are not solicited by the meaning of any word contained in the sentence. Here are just a few examples. The aspect of what is said determined by context but not via the meaning of any word appears between brackets.

It's raining *[in Paris]*.
Every student *[in my group]* failed the exam.
The queen has arrived *[at the palace]*.
I'm ready *[to leave the house]*.
I've had breakfast *[this morning]*.
The apple is green *[on the outside]*.
Mum and I went to Venice *[together]*.

Starting with Dan Sperber and Deirdre Wilson, several theorists have presented cases of this sort as counterexamples to *semanticity* (Bach 1994b, Carston 2002, Recanati 2003, Sperber & Wilson 1995). Indeed, the lesson of these examples seems to be that some aspects of what is said are not determined by the syntax and meaning of the uttered sentence. In response, Jason Stanley has defended *semanticity* by contending that all the sentences above possess more constituents than they possess words. Below the level of surface grammar, one finds unpronounced constituents at the deeper level of *logical form*. And the contextual features that determine the aspects of what is said between brackets are solicited by these unpronounced constituents (Stanley 2007). Such a defence requires a further amendment to *semanticity*: what is said with the utterance of a sentence is fully determined by (i) the meaning of the constituents of the logical form of the sentence; (ii) rules of

meaning-composition; and (iii) *the aspects of context of utterance invoked by the meaning of the constituents of the logical form of the sentence.* Stanley's defence of *semanticity* requires strenuous work, as it demands that one give evidence of the existence and nature of unpronounced constituents for every plausible counterexample to *semanticity* - see (Elbourne 2020) for such work on 'It's raining'.

The first part of this dissertation is devoted to challenging *semanticity*, but not with the kind of examples just mentioned. That is, I do not challenge semanticity on the grounds that some aspects of what is said with the utterance of a sentence correspond to no words in the sentence. I wish to fight semanticity on another front: that of what speakers say by uttering sentences containing certain context-sensitive expressions. I concentrate on the demonstrative pronouns 'This' and 'That', sometimes extending my discussion to personal pronouns such as 'He' and 'She'. Intuitively, while the meaning of 'I' and 'Today' gives us the contribution of these words to what is said given the context of utterance, the meaning of demonstrative and personal pronouns gives us less. The meaning of 'He' only tells us that an utterance of 'He' refers to some male being in the context of utterance. The meaning of 'This' only tells us that an utterance of 'This' refers to something relatively close to the speaker in the context of utterance.

The mainstream treatment of pronouns in contemporary semantic theory seems to record a difference in meaning between e.g. 'He' and 'I', while keeping the contribution of 'He' to what is said in the province of semantic theory through technical tricks (Heim & Kratzer 1998 chapter 9, Nowak 2019). However, if it is correct to construe the meaning of 'This' and 'He' along the lines I drew, then 'This' and 'He' are unlike 'I' and 'Today' in that their meaning does not *determine* the contribution to what is said given a context of utterance, but rather merely *restricts* it. The meaning of 'I' tells us that the contribution of an utterance of 'I' to what is said is the speaker. The meaning of 'He' tells us no such thing - that the contribution of 'He' is *the F* in the context of utterance. The meaning of 'He' merely restricts the contribution of its utterance to male beings.

David Kaplan, whose pioneering work integrated pre-theoretical insights about 'I' and 'Today' into a truth-conditional theory of meaning, was somewhat reluctant to treat context-sensitive expressions such as 'This' and 'He' in the same way as 'I' and 'Today' (Kaplan 1989a). In spite of this reluctance, a whole literature has subsequently been dedicated to providing a meaning for 'This' and 'He' that determines their contribution to what is said. This theoretical project is known as the *metasemantics for demonstratives* project (Nowak & Michaelson 2021). Officially, the question that guides this project is: what makes it the case that an utterance of 'This' or 'He' refers to a certain object? But in the background one often finds the assumption that an answer to this question puts 'This' and 'He' on a par with 'I' and 'Today' (King 2014, Speaks 2017).

The first part of my dissertation is a direct attack on the metasemantics for demonstratives project that serves as an indirect attack on *semanticity*. I take the failure of the metasemantics for demonstratives project to inflict a defeat on the advocates of *semanticity*: what is said

with a sentence containing 'This' or 'He' is not fully determined by (i) the meaning of the constituents of the sentence; (ii) rules of meaning-composition; and (iii) the aspects of context of utterance invoked by the meaning of the constituents.

The second part of the dissertation investigates how certain properties of what is said can be explained if *semanticity* is false. The property that interests me most is the *publicity* of what is said. Publicity is the property of being independent from both the speaker's intentions in producing her utterance and the audience's actual interpretation of the utterance. That what is said is public puts advocates of *semanticity* at an advantage. First, the publicity of what is said is easily explained if what is said is fully determined by syntax and meaning. Secondly, opponents of semanticity usually take what is said to be partly determined by the speaker's communicative intentions, and this view is straightforwardly inconsistent with the publicity of what is said. Opponents of semanticity are aware of this problem. Those who address it present themselves as putting forward a revisionary notion of *what is said* from which the publicity of the pre-theoretical notion has been expunged (Neale 2004, Perry 2009). Revisionary notions might be useful for various theoretical purposes, but my purpose here is to explain the publicity associated with the *pre-theoretical* notion of what is said. The challenge I set myself in the second part of the dissertation is then to provide a non-semantic explanation of the publicity of what is said. In a slogan: publicity without semanticity.

The first four chapters of the dissertation are devoted to a critique of the metasemantics for demonstratives project. In the first three chapters, I proceed indirectly by arguing against one popular family of metasemantics for demonstratives called *intentionalism*. According to intentionalism, the reference of demonstratives is determined by the intentions of speakers. In Chapter 1, I discuss the problem of predictive adequacy for intentionalism: intentionalism struggles to make intuitively correct predictions about reference in a number of cases. I consider two possible ways to deal with this problem, and argue that they both fail. In Chapter 2, I turn to conceptual problems for intentionalism. There I argue that several objections to reference-determining communicative intentions raised in the literature are unconvincing, but that the objection from redundancy - that intentionalism posits unnecessary linguistic meanings - has real bite. In Chapter 3, I zoom in on a recently developed version of intentionalism according to which reference-determining intentions are intentions about the reference relation they determine. I discuss two problems for this view: that of finding a plausible content for such intentions, and that of justifying the postulation of such intentions. I argue that the second problem sticks, partly because justifications found in the literature do not withstand scrutiny. In Chapter 4, I turn my back on intentionalism to look at the *metasemantics for demonstratives* project more generally. I introduce three desiderata for a metasemantics for demonstratives and argue that two of them - making predictions that match pre-theoretical 'say that' reports and providing a linguistic meaning for demonstratives - are unlikely to be jointly satisfied. I further explore the consequences of dropping each of the three desiderata in turn. One result of this exploration is that a theory accounting for 'say that' reports but failing to provide a linguistic meaning for demonstratives is worth having, although the reference relation that figures in such a theory is markedly different from the one usually assumed in the metasemantics project.

The second part of my dissertation consists of three chapters. In Chapter 5, I introduce two properties of what is said: *publicity* and *productivity*. These properties are easily explainable by the *semanticity* of what is said. However, I argue that some uncontroversially non-semantic contents of utterances - such as what is implied (suggested, insinuated) - also possess these two properties. This suggests that the publicity and productivity of what is said are not best explained by it being *semantic*. In Chapter 6, I take a small detour in order to be able to provide a non-semantic explanation of the publicity of what is said. I introduce the notion of a *defective content-act* and put forward a modal analysis of this defect: the content-act performed by a speaker in making her utterance is defective just when a common-ground-defined audience would not retrieve the speaker's intended content. In Chapter 7, I suggest that this modal condition can serve to predict what the speaker says: what a speaker says is what a common-ground-defined audience would take the speaker to intend. Similar modal conditions can serve to predict what a speaker implies and other speech act contents. I contend that these modal conditions track the contents speakers are responsible for, and that our ordinary practice of reporting speech indirectly can be understood as that of tracking what speakers make themselves responsible for by producing utterances.

Chapter 1: The predictive problem for intentionalism

What makes it the case that a demonstrative pronoun ('This', 'That') refers to a certain object when uttered by a speaker? A metasemantics for demonstratives is a theory that answers this question. Before investigating particular answers, let me add two preliminary qualifications about the relation of *reference* that figures in the question. As I use 'refer', an uttered demonstrative refers to an object just when its compositional contribution determines this object. This is consistent with, but does not require a referential semantics for demonstratives - i.e. the view that the compositional contribution of the demonstrative *is* the object. Alternative views about the compositional contribution of demonstratives are consistent with their referring in the sense just suggested. Their compositional contribution can be a function from variable assignments to objects (Rabern 2012, Nowak 2019). It can also be a higher-order property of the sort expressed by definite descriptions (King 2001).[1]

As for the second qualification, there are two views of the *relata* of the reference relation. One view holds that reference holds between expression-types, objects, and contexts: an expression-type refers to an object in a context (Kaplan 1989a, Braun 1996). The other view holds that reference holds between utterances of expression-types and objects (García-Carpintero 1998, Perry 1997). The problems discussed in this dissertation do not turn on the exact nature of the *relata* of reference. For convenience and consistency, I will use formulations that presuppose the expression-type view, such as: *expression e refers to object o in context c*.

I can now turn to particular answers to the metasemantic question. *Intentionalism* is a broad family of metasemantics for demonstratives. Here is a possible formulation of intentionalism:

> Demonstrative *d* refers to object *o* in context *c* only if the speaker in *c* intends to refer to *o* with *d*.

There is some controversy regarding the nature of the intention on the right-hand side of this conditional: is it a Grice-style communicative intention (Neale & Schiffer 2020); or is it an intention about the relation of reference that said intention determines (Viebahn 2020)? The formulation above is meant to be neutral between these two conceptions of the intention. A further issue is that of the sufficiency of the intention. Some authors take it to be the sole determinant of reference (Åkerman 2009, Bach 1992, Kaplan 1989b, Perry 2009), while some do not and further disagree about additional determinants (King 2014, Reimer 1992, Speaks 2016).

---

[1] In the first case, the referent is the value of the variable standing for the demonstrative in the logical form of the sentence, given the variable-assignment determined by the context of utterance. In the second case, the referent is the object that satisfies the property. To make their neutrality about the compositional contribution of demonstratives explicit, authors such as King and Speaks use 'semantic value' instead of 'reference' (King 2014, Speaks 2016).

Intentionalism is probably the most debated view in the metasemantics for demonstratives literature. It is also a popular view. In addition to the authors already cited, recent advocates of intentionalism include: Gomez-Torrente (2019), Lewerentz & Marschall (2018), Montminy (2010), Radulescu (2018), and Stokke (2010). Of course, the view does not command universal agreement. The staunchest opponents of intentionalism are probably Devitt (2022), Gauker (2008, 2018), and Stojnic (Stojnic et al. 2013).

Intentionalism has been saddled with many problems, but we can make a basic distinction between two types of problems for intentionalism - and for any metasemantics more generally. A metasemantics for demonstratives is a theory of the determinants of the reference of demonstratives. It thus makes predictions about reference in particular cases. One way in which such a theory can fail is then that it makes the wrong predictions about reference in some cases. We can speak of a *predictive* problem. The other way in which a metasemantics can fail is that its proposed determinants of reference are not plausible regardless of predictive adequacy. This implausibility may have to do with the nature of the proposed determinants, with the nature of the reference relation, with the interaction between the two, etc. We may call such problems *conceptual*.

In this chapter, I focus on intentionalism's predictive problem. Intentionalism seems to make incorrect predictions about reference in a number of cases. The first response to this problem is to explain away these predictions, by explaining why a clash between pre-theoretical judgements and theoretical predictions in certain cases does not threaten the theory. This response is not specific to intentionalism: any metasemantics facing a predictive problem can be defended in this way. I delay a general discussion of this sort of response to Chapter 4. In this chapter, I merely argue that intentionalists who have responded in this way to intentionalism's predictive problem have not done so convincingly. Now, in contrast with the first response, the second response to intentionalism's predictive problem is specific to intentionalism. It consists in maintaining that intentionalism makes the right predictions about reference once we get clear on which intention of the speaker determines reference. According to this response, the intention that determines reference occupies a certain place - the *proximal* one - in a structure of intentions. This form of intentionalism may be called *proximal intentions (PI) intentionalism.* I will argue that PI intentionalism is no solution to the predictive problem. Notwithstanding appearances, PI intentionalism suffers as much as less sophisticated forms of intentionalism from predictive inadequacy.

1. The predictive problem

Intentionalism is challenged by cases in which, intuitively, the referent is a certain object although the speaker does not intend to refer to this object. Here is the most notorious case of this sort, courtesy of David Kaplan:

> Suppose that without turning and looking I point to the place on my wall which has long been occupied by a picture of Rudolf Carnap and I say:

(27) That is a picture of one of the greatest philosophers of the twentieth century.

But unbeknownst to me, someone has replaced my picture of Carnap with one of Spiro Agnew (adapted from Kaplan 1978, p.239).

Similar cases are put forward in (Reimer 1991) and (Wettstein 1984). It is easy to generate *prima facie* counterexamples to intentionalism in the mould of Kaplan's case: make the speaker point at another object than the one she has in mind, for whatever reason. This is not a failsafe recipe, but it generates a truckload of *prima facie* counterexamples.

Opponents of intentionalism may present the following argument on the basis of the Carnap-Agnew case. The speaker intends to refer to Carnap's picture, not Agnew's picture. Yet the referent of the demonstrative is Agnew's picture. Therefore intentionalism is false. The first premise of the argument is easy enough to justify: the speaker intends to communicate a certain thought, and this thought is about Carnap's picture (Bach 1994a, pp.50-53). As for the second premise of the argument, Kaplan comments on his own case: *"I have said of a picture of Spiro Agnew that it pictures one of the greatest philosophers of the twentieth century"* (Kaplan 1978, p.239).[2] Even if we share Kaplan's judgement about what the speaker said (I do), there is a missing link between this judgement and the second premise of the argument. What is the relation between pre-theoretical judgements about what the speaker said and claims about reference? In the background of this argument, and more generally in the background of the metasemantics debate, we find the following assumption: pre-theoretical judgements about what is said track the relation of reference (the very relation whose determinants puzzle metasemanticists). I will not question this assumption in this chapter (I will do so in chapter 4).

2. Explaining away seemingly incorrect predictions

The first response to the argument from Carnap-Agnew-style cases is to deny that pre-theoretical judgements on what is said - 'say-judgements' - in these cases track reference. Discussing an apparent counterexample to intentionalism, Speaks writes:

> … it is at least reasonable to wonder whether our sense that something has gone wrong [...] is tracking [...] anything about the semantic values of the demonstratives or truth conditions of the utterance. (Speaks 2016, p.329)

Here is Radulescu on the same case:

[2] Another kind of pre-theoretical judgement that is sometimes invoked is truth-value judgement - e.g. that the speaker spoke truly or falsely. Truth-value judgements seem to depend on say-judgements: e.g. the speaker spoke falsely because she said that *p* and *p* is false.

> I share [this] unease about the case: something clearly went wrong. But my intuitions are not fine enough to trace the effect down to the reference level. (Radulescu 2019, p.779)

Neither Speaks nor Radulescu go much further in their explanation of our judgements in the problematic case. Speaks only suggests that in problematic cases for intentionalism, these judgements might be tracking *"irresponsibility on the part of the speaker"* (2016, p.329). This is an interesting suggestion, but it is not further developed. As for Radulescu, he seems to think that offering a positive theory of reference and communication dispenses one of saying more about the relation between pre-theoretical judgements on cases and the reference relation (Radulescu 2019, pp.779-780). This seems to me wrong on several counts. First, non-intentionalists too have a plausible picture of reference to offer in support of their preferred metasemantics - e.g. (Gauker 2008) and (Devitt 2022). The fact that intentionalists have their own picture to offer should not convince anyone to turn a blind eye to *prima facie* counterexamples to intentionalism. Secondly, offering a theory of reference and offering a theory of the relationship between pre-theoretical say-judgements and reference are independent theoretical tasks. Doing the former does not dispense one of doing the latter. Finally, one *should* offer some kind of view about the relationship between pre-theoretical say-judgements and reference if one casts doubt on judgements in particular cases. If it is suggested that some say-judgements do not track reference, the background assumption that say-judgements track reference must at least be qualified.

Once doubt is thrown upon some pre-theoretical say-judgements, one faces an immediate choice: do pre-theoretical say-judgements not track reference *in the problematic cases only*, or do they *never* track reference? In chapter 4 of this dissertation, I will explore each of these two options. Now, some intentionalists have attempted to justify why pre-theoretical judgements on troublesome cases are not reliable guides to reference, and so I will discuss their efforts here.

Gomez-Torrente claims that the reference of a demonstrative in a normal context is the speaker's intended referent, a normal context being one in which the speaker does not have referential intentions targeting different objects - *conflicting* referential intentions (Gomez-Torrente 2019, pp.19-59). This metasemantics is also supposed to be encoded in the linguistic meaning of demonstratives. The metasemantics is however silent about abnormal contexts of utterance, i.e. those involving conflicting referential intentions. The consequence of this silence is that our language does not tell us which object to associate with a demonstrative in cases of conflicting referential intentions. Assuming that pre-theoretical say-judgements track reference, something special should be going on with judgements on cases of conflicting referential intentions. According to Gomez-Torrente, what happens in such cases is that we try to associate an object with the demonstrative despite the silence of our language. Because of this silence, we do not get a clear verdict on what the speaker said.

Gomez-Torrente further contends that there is evidence of this unclarity: ordinary speakers allegedly struggle to come up with a clear verdict on cases of conflicting referential intentions (Gomez-Torrente 2019, pp.40-41).

Putting aside the analysis of Carnap-Agnew-style cases as involving conflicting referential intentions, is it right that our pre-theoretical say-judgements on these cases are unclear? Gomez-Torrente confesses that his own judgement is unclear. Mine is clear - the speaker said something about Agnew's picture - but Gomez-Torrente's claim of unclear judgements may perhaps be justified by the *inter-personal variability* of judgements on Carnap-Agnew-style cases. Unfortunately, the only empirical study that has been conducted on this topic does not support Gomez-Torrente's contention (Rostworowski et al. 2022).[3] Some of the cases presented to the study participants would be classified by Gomez-Torrente as cases of conflicting referential intentions - e.g. a speaker uttering 'That F', pointing in the direction of another object than the one she has in mind. Now, the overwhelming majority of the study participants judged that in such cases the speaker says something about the object being pointed at, rather than the object the speaker has in mind. Participants not only made a definite judgement although they were given a 'cannot say' option, but they also reported high confidence in their judgement. Of course, the usual reserve regarding the reproducibility of empirical results applies. But this study does suggest that the empirically testable aspect of Gomez-Torrente's view is unlikely to withstand scrutiny. And without empirical backup, Gomez-Torrente's claims about pre-theoretical judgements as well as his metasemantics are sophisticated *ad hoc* responses to apparent counterexamples to intentionalism.

Perry is another intentionalist who tries to explain away the pre-theoretical judgement that the speaker says something about Agnew's picture in the Carnap-Agnew case (Perry 2009). Perry accepts that pre-theoretical say-judgements track the relation of reference, and that this relation has an intentionalist metasemantics which is encoded in the meaning of demonstratives. But unlike Gomez-Torrente, the metasemantics includes no restriction to 'normal' contexts of utterance. So how can Perry explain away the seemingly incorrect predictions of intentionalism in the Carnap-Agnew case? Here is the twist. Perry contends that pre-theoretical say-judgements do not track the relation of reference *only*. Say-judgements track two relations at the same time, the other relation being determined by something else than the speaker's intention:

> …there is a forensic element to our ordinary concept of what is said. Saying something is often a social act, which has effects on others in virtue of the words used, their meanings, and other publicly observable indications of the speaker's intentions [...]. One is held responsible for some of these effects. (Perry 2009, p.191)

[3] I should note that this study asks ordinary speakers to make truth-value judgements (did the speaker say something true or false?) rather than say-judgements (what did the speaker say?), and that the cases presented to the subjects involved the utterance of complex demonstratives (e.g. 'That F') rather than bare demonstratives (e.g. 'That').

According to Perry, we track two distinct relations at the same time with say-judgements: one 'forensic' relation determined by the likely effect of the speaker's utterance on her audience, and one relation determined by the speaker's intentions. Only the latter has its determinants encoded in the meaning of demonstratives. In most cases the object *relatum* of the two relations is the same, and so the double-tracking goes unnoticed. But in Carnap-Agnew-style cases, the object *relatum* of the forensic relation and the object *relatum* of the intentionalist relation come apart. Perry tentatively claims that in such cases the forensic relation trumps the intentionalist relation: this is why we tend to judge that Kaplan said something about Agnew's picture (Perry 2009, p.192).

The first problem with this picture is obvious: why should one relation rather than the other be the one determined by meaning via linguistically encoded determinants? Once the theorist claims that say-judgements track several relations, which one she decides to make the linguistically determined relation seems an arbitrary choice. That is, unless there is some independent way to assess which relation is the linguistically determined one. For example, there could be a distinctive phenomenology involved in associating an object with an expression by following a linguistic rule. This hypothesis is however implausible: if there was such a distinctive phenomenology, it would be obvious e.g. which of Carnap's picture or Agnew's picture is churned out by the linguistic rule. And it would be obvious whether Perry is right or wrong to make the intentionalist relation, rather than the 'forensic' relation, the linguistically determined one.

The second problem is perhaps more subtle. Given the results of the empirical study discussed earlier, Perry's view seems to be empirically adequate. However, this adequacy comes at a cost. If I understand Perry correctly, our say-judgements always track the 'forensic' content. It's just that outside of Carnap-Agnew-style cases the forensic content is the same as the intentionalist content, and so the *double*-tracking aspect of say-judgements remains concealed. But then Perry's view is empirically indistinguishable from the view that say-judgements always track forensic content *only*. Considerations of economy and explanatory simplicity should steer the theorist towards the latter view. This point can be put in more general terms. Claiming all at once (i) that say-judgements track two different relations R1 and R2, (ii) that R1 is the relation of reference, (ii) but that R2 prevails when there is a conflict between R1 and R2 has the self-defeating consequence of supporting a view on which say-judgements track R2 only and R2 is the relation of reference.

I do not know if Perry anticipates this worry, but he tries to provide evidence of double-tracking. Perry claims that in Carnap-Agnew-style cases our say-judgements display a distinctive variability. This variability is *intra-personal*: depending on certain features of the audience of the case, the same observer will be tempted to judge that the speaker said something about Agnew's picture or that she said something about Carnap's picture (Perry 2009, pp.191-192). This is true even if the forensic content tends to be more salient. This kind of intra-personal variability is not tested by the empirical study already cited, as this study does not offer several variants of each case, with different audience-features. Granting for the sake of argument that Perry is right about this form of intra-personal variability, his

explanation is again not the most simple explanation. Here is the most simple explanation: say-judgements track only the forensic relation, and the forensic relation is partly determined by features of the audience. Compare this with Perry's explanation. Say-judgements track two relations: the forensic one and the intentionalist one. In cases where the object *relata* of these relations come apart, it is possible to make one relation or the other more salient by tinkering with the features of the audience.

There is more to say about the different ways of questioning the assumption that pre-theoretical say-judgements track reference. In this section I have only shown that intentionalists who respond to Carnap-Agnew-style cases by questioning our pre-theoretical judgements have not done enough to convince us that these cases should not be treated as counterexamples to intentionalism.

3. Proximal intentions (PI) intentionalism

a. Introducing PI intentionalism

Here is another response to the argument against intentionalism from the Carnap-Agnew case. The argument mistakenly assumes that the speaker of the Carnap-Agnew case intends to refer to Carnap's picture only. In fact, the speaker intends to refer to Agnew's picture too. This further intention is usually presented as an intention to refer to *the F*, where the description 'the F' is satisfied by Agnew's picture. For example, it seems true that the speaker intends to refer to *the picture behind him*. The second step of the response consists in claiming that the intention about Agnew's picture prevails in the determination of reference. This response to the Carnap-Agnew case may be called the 'multiple intentions' response.

The multiple intentions response is unfortunately *ad hoc*. We are conveniently told that the intention about the intuitive referent is the reference-determining one, but no justification is given. The situation is actually worse. The intention *to refer to the picture behind her* can be ascribed to the speaker on the following basis: she intends to refer to Carnap's picture and believes that Carnap's picture is *the picture behind her*. But if we follow this rationale, every belief (or at least every explicit belief) of the speaker that Carnap's picture *is the F* comes with a corresponding intention to refer to *the F*. The question is then not just why the reference-determining intention is the one about the picture behind the speaker rather than the one about Carnap's picture. One must also explain why it is not the intention about *the F.*

What we need is a general theory that predicts the ascendancy of the intention about the intuitive referent in the Carnap-Agnew case. This is where PI intentionalism enters the stage. Several intentionalists go beyond the multiple intentions response (Bach 1992, King 2013, Perry 2009, Reimer 1992). They first observe that the intention about Carnap's picture and the intention about Agnew's picture are part of a common structure of intentions, and that the intention about Agnew's picture occupies a certain place - the *proximal* one - in this structure.

They further propose that only *proximal intentions* determine reference. This is the view I call PI intentionalism.

Let me sketch what I take to be the best version of PI intentionalism. This version anchors itself to a general view about intentional action. Among PI intentionalists, only King ties his metasemantics to a general view of this sort (King 2013).[4] One might call this general view the doctrine of structured intentions. The doctrine of structured intentions is widely endorsed, not only in philosophy (Bratman 1990, Mele 1992, Searle 1980) but also at the border of philosophy and psychology (Pacherie 2008).

One starting point for the doctrine of structured intentions is the platitude that we often do something by doing something else. I volunteer by raising my hand. I score a try by grounding the ball behind the line. These descriptions of actions reflect the further platitude that we attain a certain end by employing certain means. According to the doctrine of structured intentions, if an action is fully intentional, then to each level of description of an action corresponds some intention. When I intentionally score a try by grounding the ball, I have an intention corresponding to the 'score a try' level of description, and I have an intention corresponding to the 'ground the ball' level of description. Furthermore, an explanatory structure ties these intentions together: I intend to ground the ball *because* I intend to score a try.[5] One may call the explanatory intention the *distal* intention, and call the explained intention the *proximal* intention.[6] [7] Structured intentions are sometimes captured through formulations such as 'A intends to φ by ψ-ing' - e.g. the student intends to volunteer by raising her hand. The distal intention attaches to φ, and the proximal intention attaches to ψ.

I will sometimes speak of proximal and distal *acts*, but this is just for convenience: a metaphysics of structured acts is not strictly required. Neither need we assume that the intention corresponding to a level of description of the action φ is the intention to φ - what Bratman calls "the simple view" (Bratman 1984). All we need is distinct intentions corresponding to different levels of description of an action, and these intentions to be ordered by an explanatory relation.

Here is how PI intentionalists may apply the doctrine of structured intentions to utterances, and more specifically to the Carnap-Agnew case. In the Carnap-Agnew case, the action of the speaker may be described as follows: she expresses a thought about Carnap's picture by pointing at the picture behind her and uttering 'That is a picture…'. The speaker's distal

[4] King specifically appeals to Bratman's theory of intentions (Bratman 1990). Bach ties his version of PI intentionalism to the more local view that *communicative* intentions are structured (Bach 1992).
[5] Bertolet and Wettstein attribute this idea of a hierarchy of intentions related by explanatory relations to Castañeda (Bertolet 1987, Castañeda 1971, Wettstein 1984).
[6] I borrow the 'proximal/distal' terminology from Pacherie, who borrows it from Mele (Mele 1992, Pacherie 2008). One finds great terminological variety in the metasemantics literature.
[7] There might be more than two component intentions involved, in which case 'proximal/distal' should be interpreted as a continuum rather than a discrete binary distinction. For present purposes, a binary distinction will do.

intention attaches to the level of description of the action before 'by'. PI intentionalists claim that this intention picks out Carnap's picture. The speaker's proximal intention attaches to the level of description of the action after 'by'. PI intentionalists claim that this intention picks out the picture behind the speaker, i.e. Agnew's picture.

Once anchored to the doctrine of structured intentions, PI intentionalism seems to offer the most principled intentionalist response to the Carnap-Agnew case. First, the ascendancy of the intention about Agnew's picture is motivated by the general ascendancy of proximal intentions. Secondly, the proximal status granted to the intention about Agnew's picture is motivated by a general view connecting descriptions of action and intentions. I take this 'anchored' version of PI intentionalism as my target at the start of the next section. I take on 'non-anchored' versions of PI intentionalism later in section 3.d.

b. Ostensive proximal intentions

Let us consider the Carnap-Agnew case once again. Since the speaker intentionally points at the picture behind her, it is intuitively the proximal intention attaching to her pointing gesture that secures the right prediction for PI intentionalism. Let us call proximal intentions attaching to ostensive gestures *ostensive intentions*. If the PI intentionalist's take on the Carnap-Agnew case is correct, the ostensive intention of the speaker is about the picture behind her. This ostensive intention is determinate, in the sense that it picks out a unique object. Now, there is a tension between this take and the widely acknowledged view that ostensive *gestures* are indeterminate (Kaplan 1989a, King 2014, Reimer 1992). The Carnap-Agnew case is a case in point: the speaker's ostensive gesture does not determine Agnew's picture more than e.g. its frame, the nail on which the frame hangs, or the glass screen protecting the picture.

A gap between indeterminate ostensive gestures and determinate ostensive intentions needs to be filled. At this point it seems natural to let the speaker's beliefs fill the gap. The speaker wants to communicate a thought about Carnap's picture, and she has beliefs of the form *Carnap's picture is the F*. It seems then right to attribute to the speaker the ostensive intention e.g. to point at *the F*. The speaker believes that Carnap's picture is the picture on the wall behind her, and so her ostensive intention is to point at the picture on the wall behind her.

This line of thought faces the immediate problem that the speaker might have several beliefs of the form *Carnap's picture is the F*, and that some of these beliefs might not target Agnew's picture. Sure, the speaker believes that Carnap's picture is *the picture on the wall behind her.* But she might also believe (truly let's say) that Carnap's picture is *her ten-year anniversary present*. If the former belief fixes the content of the ostensive intention, PI intentionalism makes the right prediction about reference. But if the latter belief does, PI intentionalism makes the wrong prediction. In addition, one cannot arbitrarily stipulate that the former belief trumps the latter belief. Relying on descriptive beliefs to make the content of ostensive intentions determinate thus leads to problems similar to those afflicting descriptivist

metasemantics for proper names. This has been noted by Speaks (2017, p.731) and Devitt (2022, pp.1000-1001). A double threat of misdescription and arbitrariness looms: some of the speaker's beliefs denote another object than the intuitive referent, and one cannot arbitrarily stipulate that the beliefs denoting the intuitive referent are the content-fixing ones.[8]

There is an intuitive way out of this problem. The speaker's belief that Carnap's picture is the picture on the wall behind her is intuitively relevant to the pointing gesture to which her ostensive intention attaches. By contrast, her belief that Carnap's picture is her ten-year anniversary present is intuitively irrelevant to her pointing gesture. Can we make good on this intuitive contrast? Reimer writes: "*The relevant beliefs will be those that connect the intended demonstratum (the object of the primary* [i.e. distal] *intention*) *with the demonstrative act*" (Reimer 1992, p.390). Reimer does not say what the nature of this connection is, but one can extract from King's work the idea that the connection is explanatory (King 2013). The beliefs that fix the content of ostensive intentions are those that in some intuitive sense explain the speaker's ostensive act. In the Carnap-Agnew case, the speaker points as she does *because* she believes that Carnap's picture is the picture on the wall behind her. This *explanatory belief* fixes the content of the ostensive intention, which turns out to be determinate and to pick out Agnew's picture. Or so the story goes.

There is something wrong with this story. Why does the speaker of the Carnap-Agnew case make the ostensive gesture that she makes? She points behind her because she believes that Carnap's picture is *on the wall behind her*. The explanatory belief of her ostensive act is then really a belief about the location of Carnap's picture. And this location belief does not target Agnew's picture more than its frame, than the nail on which it hangs, etc. Sure, Agnew's picture is *on the wall behind the speaker*. But so are the frame, the nail, etc. Explanatory beliefs of ostensive acts cannot buy us determinate ostensive intentions.

The story presented two paragraphs ago tries to conceal this by smuggling additional properties into the explanatory belief - e.g. *the F* on the wall behind the speaker. The reality is that beliefs other than the explanatory belief must be recruited to yield an ostensive intention to point at *the F* on the wall behind the speaker. And this brings us back to the double threat of misdescription and arbitrariness. Suppose that the speaker believes that Carnap's picture is a painting, and further believes that it is the painting on the wall behind her - call this further belief B1. Now suppose that Agnew's picture is not a painting, but a photograph. If belief B1 is allowed to fix the content of the speaker's ostensive intention, this ostensive intention does not pick out Agnew's picture. That's the misdescription problem. Of course, the speaker also believes that Carnap's picture is the *picture* behind her - call this belief B2. But why should B2 (rather than B1) fix the content of the speaker's ostensive intention? That's the arbitrariness problem.

[8] Speaks takes the problem to be slightly different. He sees here a double threat of misdescription and *psychological implausibility* (Speaks 2017). With psychological implausibility, the idea is that the description which would secure the right prediction about reference cannot play a role in the explanation of the speaker's action.

We have tried to escape the problem of misdescription and arbitrariness for ostensive intentions by appealing to explanatory beliefs, but this appeal has only led us back to it. Here is the problem in its most general form: if the gap between indeterminate ostensive acts and determinate ostensive intentions is filled by descriptive beliefs, then PI intentionalism faces a double threat of misdescription and arbitrariness. Now, one can try to reject the antecedent of this conditional on various grounds. One might first deny that the gap between indeterminate ostensive acts and determinate ostensive intentions is filled by beliefs. The alternative is a form of externalism according to which the content of the ostensive intention is fixed by facts beyond the speaker's mental states. I do not know what form such a view could take. I myself cannot think of an externalist mechanism of determination of intention-content which is both independently plausible and guarantees that Agnew's picture is picked out by the speaker's ostensive intention in the Carnap-Agnew case.

Another option is to grant that beliefs fix the content of ostensive intentions while denying that content-fixing beliefs must be *descriptive*. Content-fixing beliefs could instead be fully *de re*, e.g. believing of Carnap's picture and Agnew's picture that they are identical. This might then yield a *de re* ostensive intention about Agnew's picture - e.g. to point at *it.* I see no reason to bar *de re* beliefs from fixing the content of ostensive intentions in general. PI intentionalists are open to this too (King 2013, p.301; Perry 2009, p.190; Reimer 1992, pp.391-392). However, the local consensus on the Carnap-Agnew case seems to be that the speaker cannot have mental states whose content includes Agnew's picture itself. This is presumably because the speaker has never seen Agnew's picture, and has never even heard of it. None of the relations between thinker and object that are usually regarded as allowing *de re* thought (perception, memory, communicative chains) holds between the speaker and Agnew's picture.[9]

Now, there are several views in the literature that allow *de re* thought in the absence of such relations. But even among these more liberal views, many do *not* predict that in the Carnap-Agnew case the speaker thinks *de re* about Agnew's picture. To give just one example, Jeshion proposes that an object being significant to an agent, in the sense that it has a considerable impact on her cognitive and affective life, is enough for the agent to think about this object *de re* (Jeshion 2010). For instance, an adoptee who fervently hopes that she will one day meet her unknown biological mother thinks of her *de re*. In the Carnap-Agnew case, Agnew's picture is not significant to the speaker in the relevant sense, and so Jeshion's liberal view does not allow the speaker to think of Agnew's picture *de re*. As far as I can see, the only theory of *de re* thought that could allow this is one according to which agents can voluntarily introduce a name-like mental vehicle to think about an object satisfying a certain description, and this voluntary introduction is enough to make the object enter the content of their thought.[10]

[9] For a classic statement of this view of *de re* thought, see part one of (Bach 1994a).
[10] This view is associated with Harman (1977). It has few contemporary advocates: as far as I know, only Sainsbury endorses it (Sainsbury 2020).

Even if this hyper-liberal view is endorsed, it is far from clear that a correct prediction on the Carnap-Agnew case can be reached. Assuming that the speaker can think of Agnew's picture *de re*, she suffers from confusion: she takes Carnap's picture and Agnew's picture to be one and the same object.[11] Different positions have been taken on the content of the states of agents who suffer from object-confusion, but none is that only one of the confused objects is the content. One view is that confused *de re* thoughts are empty (Lawlor 2007, Recanati 2012).[12] Another is that they partially refer to each of the confused objects (Recanati 2016).[13] Yet another view is that they have a non-actual object as their content (Milikan 2000, Unnsteinsson 2019). Overall, PI intentionalists are not in a position to claim that the speaker of the Carnap-Agnew case has a *de re* ostensive intention about Agnew's picture. More generally, I conclude that PI intentionalists are not in a position to claim that the speaker has a determinate ostensive intention - either descriptive or *de re* - about Agnew's picture.

A concessive response to this problem is crying for consideration. In the Carnap-Agnew case, the proximal level of the speaker's action does not stop at her ostensive gesture: she also utters a sentence. The intentions attached to the linguistic part of her proximal act have so far been overlooked. One could then tentatively accept that ostensive intentions do not pick out Agnew's picture on their own, and hope that *linguistic* proximal intentions get us over the line. Considering linguistic proximal intentions also seems necessary beyond the Carnap-Agnew case. Sometimes the utterance of a demonstrative is not accompanied by an ostensive gesture. If PI intentionalism is to have any chance of accounting for the reference of demonstratives in such cases, one had better look at linguistic proximal intentions.

c. Linguistic proximal intentions

How should one conceive of *linguistic* proximal intentions? I will take an instrumental approach to this question: first set out what one would like these intentions to do for PI intentionalism, and then characterise these intentions so that they can do the expected job. Consider the Carnap-Agnew case once again. If one grants that the speaker' ostensive intention is indeterminate, one may regard its contribution to reference-determination as a mere restriction to objects that are *on the wall behind the speaker*. One would then like the content of the speaker's linguistic intention to include e.g. the property of being a picture, so that the overall proximal intention restricts reference to an object that is (i) on the wall behind her and (ii) a picture. This would be enough for PI intentionalism to secure the right prediction about reference. In general, linguistic proximal intentions should be about properties possessed by the intuitive referent of the demonstrative.

To deserve their label, linguistic intentions must be traceable to the linguistic part of the speaker's intentional action, that is, to her intentional utterance of linguistic expressions. And

[11] There seem to be two main theoretical glosses on the phenomenon of object-confusion. See (Lawlor 2005) and (Unnsteinsson 2016) for introduction and discussion.
[12] For (Recanati 2012), see chapters 10 and 11 (pp.115-144).
[13] For (Recanati 2016), see chapter 2 (pp.14-31).

since linguistic intentions should target properties of the intuitive referent, they should be e.g. intentions *to refer to a F* rather than e.g. intentions *to utter the expression 'F'*. In the Carnap-Agnew case, the speaker utters the words 'is a picture' intending them to have their conventional meaning in English. Here I assume for the sake of argument that this yields an intention *to refer to a picture*. The speaker intends to refer to a picture, and she intends to refer to something on the wall behind her. Her overall proximal intention determines Agnew's picture, or so the hope goes.

Unfortunately, this apparatus leads to a renewed double threat of misdescription and arbitrariness. In the Carnap-Agnew case, the speaker utters: 'That's a picture of one of the greatest philosophers of the twentieth century'. For the same reasons that she intends to refer to *a picture*, the speaker intends to refer to *a picture of one of the greatest philosophers of the twentieth century*. But Agnew's picture does not instantiate the latter property, and so Agnew's picture is not predicted to be the referent. Of course, one could pick and choose which part of the speaker's predication enters her linguistic intention (i.e. only 'picture' matters), but this would be arbitrary.

The introduction of reference-determining intentions attached to acts of predication has a further damning consequence. If uttering 'This/that is a F' comes with an intention to refer to a F, and if this intention determines the reference of the demonstrative in subject position, then it seems impossible to say something false of an object by uttering a sentence of the form 'This/That is F'. The possibility of saying something false of an object with a sentence containing a demonstrative in subject position depends on the possibility that the demonstrative has a referent which does not satisfy the predicate. But this possibility vanishes if the reference of the demonstrative is determined by the predicate. This consequence is unacceptable. Linguistic intentions associated with the predicative part of a sentence whose subject is a demonstrative do not contribute to the determination of reference.

Why did the contrary ever seem plausible? Well, hearers often use the predicative part of 'This/That is F' to determine the reference of the demonstrative, and speakers expect them to do so. The relevant sense of 'determine the reference' in the previous sentence is something like *ascertain*: one could call it the interpretive sense of 'determine'. The interpretive sense of 'determine reference' is distinct from its metaphysical sense, which concerns the facts in virtue of which a demonstrative refers. Only the metaphysical sense of 'determine reference' is relevant when it comes to providing a metasemantics for demonstratives. Some authors think that a confusion between the two senses of 'determine reference' besets a wide range of metasemantics for demonstratives: the facts that speakers use to determine reference in the interpretive sense are mistaken for the facts that determine reference in the metaphysical sense (Bach 2001, Neale & Schiffer 2020). This confusion might explain the misguided attempt to elevate predicative linguistic intentions to determinants of reference.

If predicative linguistic intentions really have no reference-determining power, then PI intentionalism must rely on ostensive intentions to make the right prediction about reference in the Carnap-Agnew case. I argued earlier that ostensive intentions cannot be trusted to do

the job. Beyond Carnap-Agnew-type cases, it is hard to see how PI intentionalism can account for cases in which a speaker utters 'That is F' without making any ostensive gesture and her demonstrative intuitively refers to a certain object. There is no ostensive act, and thus no ostensive intention. As for the speaker's linguistic act, the intention associated with uttering the predicate '... is F' has no reference-determining power. There are then no proximal intentions left to determine reference.

d. No way out for PI intentionalism

The version of PI intentionalism I have attacked thus far is anchored to the general doctrine of structured intentions. We have just seen that this version cannot secure some predictions about reference. Can this problem be overcome by untying PI intentionalism from the doctrine of structured intentions?

Most PI intentionalists do not tie their view to a general theory of intentional action (Bach 1992, Perry 2009, Reimer 1992). However, they offer more than an *ad hoc* response to the Carnap-Agnew case: they do not just pick a speaker-intention that happens to denote Agnew's picture and call it proximal. These PI intentionalists use instead the *prima facie* acceptability of a characterisation of structured intentions as a guide to proximal (and distal) intentions. This practice constrains the postulation of proximal intentions to some extent. Let me illustrate this point. In the Carnap-Agnew case, it seems true that the speaker intends to refer to Carnap's picture by referring to the picture behind her. And it seems false that she intends to refer to Carnap's picture by referring to e.g. her ten-year anniversary present. Hence, the intention to refer to the picture behind her is a good candidate for proximality, while the intention to refer to her ten-year anniversary present is not.

Unfortunately for this version of PI intentionalism, the *prima facie* acceptability of a characterisation of structured intentions is not discriminating enough a criterion to avoid the double threat of misdescription and arbitrariness. Say that the speaker believes that Carnap's picture is a painting. And say that Agnew's picture is in fact a photograph. The following characterisation of the speaker's structured intention seems acceptable: the speaker intends to refer to Carnap's picture by referring to the *painting* behind her. But the proximal intention yielded by this characterisation does not target Agnew's picture. So the threat of misdescription remains. Now, the following characterisation seems equally acceptable: the speaker intends to refer to Carnap's picture by referring to the *picture* behind her. And the proximal intention yielded by this characterisation targets Agnew's picture. But since the two characterisations are equally acceptable, it would be arbitrary to take the latter but not the former as a guide to the speaker's proximal intention. Arbitrariness lurks again. I conclude that untying PI intentionalism from the general doctrine of structured intentions cannot save PI intentionalism from predictive inadequacy.

Conclusion

An intentionalist metasemantics for demonstratives faces a predictive problem. It apparently makes incorrect predictions about reference in some cases (section 1). There are two main strategies to deal with the predictive problem. One can first explain away the pre-theoretical judgements that clash with theoretical predictions. But intentionalists who have gone down this path have not done so convincingly (section 2). One can alternatively move to a more sophisticated form of intentionalism, *proximal intentions* intentionalism, which apparently makes correct predictions where a less sophisticated form of intentionalism falters. This more sophisticated form of intentionalism also struggles to make correct predictions in problematic cases for intentionalism (section 3).

Chapter 2: Conceptual problems for communicative intentionalism

In the previous chapter I explored intentionalism's predictive problem. In doing so I left open the question of the nature of the intentions that determine the reference of demonstratives. A recent suggestion in the literature is that reference-determining intentions are *semantic*: they are about the very relation of reference that they determine (King 2014, Viebahn 2020). However, the default view is to regard reference-determining intentions as Gricean *communicative* intentions - e.g. (Stokke 2010). I will deal with semantic intentions in the next chapter. In the present chapter, I focus on communicative intentions. I assess potential problems other than predictive inadequacy for a form of intentionalism according to which communicative intentions determine reference - for *communicative intentionalism*.

Communicative intentions are introduced by Grice as part of his attempt to elucidate what it is for someone to mean that *p* (Grice 1989). Grice's proposals have been modified over the years (e.g. Bach & Harnish 1979, Neale 1992, Schiffer 1972), and one development of this theoretical project has been to build a notion of *speaker reference* out of the notion of speaker meaning. Speaker reference to an object *o* may be considered as a part of, or as an abstraction from the act of meaning that *p*, where *p* is a proposition that is *o*-dependent: *"S referred to x in (the course of) uttering u iff in uttering u, S meant an x-dependent proposition"* (Neale & Schiffer 2020). A full elucidation of the notion of speaker reference yields something more sophisticated, as illustrated below:

> Speaker S refers to object *o* with *e* in uttering U on a given occasion iff U is an utterance of sentence type σ, *e* is an expression type contained in σ, (∃H)(∃p) such that the truth condition of p is *o*-dependent and S utters U intending:
>
> (1) to produce thereby in H, the belief that *p*;
> (2) H to use the *e*-part of U as direct evidence that the belief S intends to produce in H is about *o*;
> (3) H to recognize S's intentions (1) and (2)
> (Unnsteinsson 2019, p.270)

The detail of the analysis of speaker reference is not crucial for the purposes of this chapter. What matters is that reference-determining communicative intentions are the intentions constitutive of speaker reference. If one holds that communicative intentions determine the reference of a certain expression *e*, one holds that the intentions constitutive of *speaker reference with an utterance of e* also determine the *reference of e in a context of utterance*. The metasemantics presented in (Neale & Schiffer 2020) makes this clear. For a range of expressions including pronouns, demonstrative pronouns, and complex demonstratives, the authors propose the following metasemantic schema:

x is the referent of expression *e* on an occasion of use *o* if and only if:
(1) C*x*
(2) the speaker refers to *x* with *e* on *o*

C is a constraint set by the meaning of expression *e*. In addition to being referred to by the speaker (*speaker reference*), *x* must satisfy this constraint in order to be referred to by *e* (*semantic reference*). For instance, the metasemantics for 'She' is:

x is the referent of 'She' on an occasion of use *o* if and only if:
(1) *x* is female
(2) the speaker refers to *x* with 'She' on *o*

Here and in other places I use 'semantic reference' to draw a contrast with speaker reference. 'Semantic' here indicates that the relation of reference in question is one between *expression and object*, as opposed to one between *speaker and object* in the case of speaker reference. It might be that such a relation between expression and object makes sense only if it is determined by the meaning of the expression, and is thus *semantic* in the sense of the introduction to this dissertation, but here the terminology 'semantic reference' is merely used to distinguish relations with different *relata*.

If one is sceptical of a notion of speaker reference defined in terms of Gricean communicative intentions, one will not favour a metasemantics that makes semantic reference dependent on speaker reference. Such scepticism might play a role in the rejection by some authors of communicative intentionalism, but here I ignore such scepticism.[14] Whether or not there is something wrong with speaker reference understood in terms of communicative intentions, one can ask if there is something wrong with the determination of semantic reference by communicative intentions. As we shall shortly see, philosophers have had many thoughts on the latter question.

Two more qualifications are needed before I can turn to objections to communicative intentionalism. First, one may question whether giving a metasemantics for an expression requires taking the linguistic meaning of the expression to encode the determinants of reference. I will question this in chapter 4, but here I assume that intentionalism about an expression commits its proponents to a meaning for the expression that encodes its intentionalist metasemantics. Secondly, I use 'reference' and 'content' interchangeably in this chapter. The content of a sub-sentential expression can be an object, a relation, or a property of any order. 'Reference' is usually employed more restrictively for *object*-contents. The

[14] For example, I ignore the complaint that Gricean communicative intentions *over-intellectualise* communication. The thought is that speakers can fully communicate without having complex Gricean intentions (Moore 2017, Thompson 2014). This objection is sometimes echoed in the metasemantics literature as an objection to intentionalism (Devitt 2022, Speaks 2016). The echoed objection goes as follows. Some unsophisticated speakers (notably children) use demonstratives, and their demonstratives refer. These unsophisticated speakers do not have sophisticated semantic or communicative intentions. Therefore intentionalism is false.

arguments in this chapter do not depend on the type of content-entity considered (object, property, or relation). This is why I use 'reference' and 'content' interchangeably.

1. Divergence between saying and meaning

Kent Bach sometimes argues as follows. If communicative intentions determine what the speaker says in addition to determining what the speaker means, then it is not possible for the speaker to say something she does not mean. And yet it happens that a speaker says something she does not mean (for example when she speaks non-literally). Therefore communicative intentions are not determinants of what the speaker says (Bach 2001 p.18, Bach 2005 p.27). Authors like King and Viebahn take this argument as one against communicative intentionalism (King 2014 p.232; Viebahn 2020 pp.96-97).[15]

This argument seems to me uncontroversial if the notion of *what is said* figuring in the first premise and the conclusion is just the ordinary, pre-theoretical one that figures in the second premise - the premise that sometimes speakers say something they do not mean. The problem is that the *what is said* of the conclusion, which Bach is eager to preserve from speaker-meaning, is also supposed to be the linguistically determined content of the uttered sentence - i.e. its *semantic* content in the sense of the introduction to this dissertation (Bach 2001). This second notion of what is said is a theoretical one. This is made particularly clear by Viebahn's own rendering of the argument, whose conclusion is that communicative intentions cannot determine the 'semantic value in context' of context-sensitive expressions (Viebahn 2020, pp.96-97). Now, the conclusion that what is said *in the pre-theoretical sense* is not determined by communicative intentions is warranted by the premise that sometimes speakers say something they do not mean. But in order to conclude that what is said *qua* linguistically determined content is not determined by communicative intentions, one must add the premise that the two notions of *what is said* coincide. In other words, one must add the premise that pre-theoretical say-judgements track linguistically determined content.

This additional premise is controversial. Relevance theorists and philosophers including Bach himself have produced a raft of examples suggesting that the linguistically determined content of a sentence is often less than a full truth-condition (Bach 1994b, Carston 2002, Sperber &

[15] King thinks that this argument does not threaten communicative intentionalism about demonstratives because the fact that sometimes speakers say something they do not mean is consistent with the referential part of what is said - e.g. the object to which a property is attributed - being determined by the referential part of communicative intentions (King 2014, p.232). It seems to me that King is wrong. At least when it comes to non-literality, divergence between what a speaker says and what she means can very well affect the referential part of what is said. When I utter 'Albert Einstein is back at it' to mock an idiotic colleague, I say something but mean something else, and intuitively this divergence finds its source in the referential part of the utterance. To be clear, I am somewhat sceptical of characterising non-literal speech as *saying something one does not mean*, but my point here is just that if this characterisation is accepted when the intuited divergence finds its source in the predicative part of the utterance, it should be equally accepted when its source is the referential part of the utterance.

Wilson 1995, Recanati 2004). Since what a speaker says in the pre-theoretical sense is truth-conditional, they conclude that pre-theoretical say-judgments track more than linguistically determined content. Other philosophers think that the linguistically determined content of a declarative sentence is always truth-conditional but comes apart from what gets reported as what the speaker says (Borg 2004, Cappelen & Lepore 2005).

Putting aside Bach's argument, what is the connection between the thesis that linguistically determined content is not determined by communicative intentions on the one hand, and the thesis that communicative intentions determine the reference of an expression on the other? Assuming that the determinants of the reference of an expression are encoded in its linguistic meaning, the expression's reference in context surely counts as a contribution to linguistically determined content. Given this assumption, the thesis that linguistically determined content is not even partly determined by communicative intentions is inconsistent with the thesis that communicative intentions determine the reference of some expression. The crucial issue for communicative intentionalism raised by Bach's argument is then whether the linguistically determined content of a sentence in context is free from the influence of communicative intentions. Some philosophers have argued that communicative intentions cannot be determinants of linguistically determined content. Let me now turn to these arguments.

2. The interpretation-dependence of communicative intentions

One family of objections to communicative intentionalism insists on the difference between communicative intentions on the one hand, and uncontroversial reference-determining contextual features - such as the identity of the speaker for the word 'I' - on the other hand. To understand these objections, it is useful to start with the distinction between *broad* and *narrow* context (Bach 2005 pp.39-40, Recanati 2004, pp.56-57). More accurately, 'broad' and 'narrow' designate two roles that non-linguistic facts - henceforth *contextual facts* - can play in linguistic communication. Contextual facts can first be invoked by the meaning of a linguistic expression. One way in which this happens is when the meaning of the expression is a rule that relates the expression to a content through a contextual fact. Such a meaning encodes the metasemantics of the relevant expression. Here is a meaning of this sort: *'I' refers in a context of utterance to the speaker in this context*.[16] Following Kaplan, such a meaning may be represented as a function from context to content (Kaplan 1989a). When a contextual fact is invoked in this way by the meaning of an expression, we may say that it plays a *narrow* role with regard to the expression. Expressions endowed with a meaning of this sort may

[16] In this chapter I am again using formulations that suggest an *expression-in-context* theory of the reference of indexicals, rather than an *utterance* theory (Braun 1996, García-Carpintero 1998). Kaplan's theory is a pure expression-in-context theory. In a pure expression-in-context theory, one cannot have the meaning of an expression invoke utterances of the expression. For example, Kaplan would describe the meaning of 'I' as: 'I' - the expression-type - refers in a context *c* to *the agent of c* (rather than to *the person who utters 'I'*). It seems to me that an intentionalist meaning should invoke *utterances*, but I will still use Kaplan as the theoretical anchor in this section. As far as I can see, the distinction between utterance theories and expression-in-context theories is not crucial for my purposes.

further be called *indexicals*. By definition, indexicals are expressions for which some contextual fact plays a narrow role.

Contextual facts can play a different role with regard to an expression. They can serve as clues for the interpreter in her attempt to recognise the speaker's communicative intention in uttering said expression. This is the *broad* role of context. For instance, the fact that a speaker makes an ostensive gesture when uttering the pronoun 'She' presumably plays a broad role: it is not invoked by the conventional meaning of 'She', but rather helps the interpreter figure out who the speaker refers to with her utterance of 'She'.[17] 'She' might still be an indexical, if some other contextual fact plays a narrow role for it.

A recurring theme in the literature is that communicative intentions are unlike established narrow contextual facts in some crucial respect, which prevents them from playing the narrow role. Let me start with a passage in (Brandom 2008):

> What I want to call 'genuine' semantic indices are [contextual] features of utterances that can be read off without knowing anything about what the utterance means. Time, place, speaker, and possible world are properties of tokenings that can be settled and specified before one turns one's attention to the content expressed by those tokenings… It [the *speakers' referent*] is not, as time, place, and speaker are, something that can be determined independently of semantic value and then appealed to as an input from which the value could then be computed… (Brandom 2008, p.58)[18]

With the terminology 'semantic indices', Brandom assumes the standard implementation of narrow contextual facts in a model-theoretic semantics: narrow contextual facts are represented as parameters in a sequence called an *index*. If communicative intentions are regarded as narrow contextual facts, then formally a parameter must be added to the index. The standard way to do this is to make *the intended referent* itself the additional parameter (Kaplan 1989a, pp.527-529).

One can draw from the quoted passage a contrast between established narrow contextual facts - e.g. who is speaking in the case of 'I' - and communicative intentions. Established narrow contextual facts can be known independently of the interpretation of the expressions whose meaning appeals to these facts. For example, it is possible to know who is speaking before engaging in the interpretation of an utterance of 'I'. Established narrow contextual features are *interpretation-independent*. By contrast, communicative intentions are *interpretation-dependent*. In order to know what a speaker means with an utterance of an

[17] Of course, some authors believe that ostensive gestures are determinants of semantic reference, although they are not always clear on whether ostensive gestures figure in the meaning of demonstratives (Wettstein 1984, Reimer 1992).

[18] Brandom is actually drawing a contrast between established narrow contextual features - speaker, time, place - and Kaplanian *demonstrations*. His point however carries over to communicative intentions, as explained by Recanati (Recanati 2010. p.3).

expression, one must antecedently engage in the interpretation of this utterance. In addition to this conceptual point, it seems that interpreters typically rely on the linguistic meaning of uttered expressions to ascertain what the speaker means by uttering them. The recognition of communicative intentions thus typically rests on antecedent semantic work on the part of interpreters.[19]

As I see it, one cannot extract an argument against communicative intentions *qua* narrow contextual facts from interpretation-dependence alone. However, there is a reading of Brandom's passage that suggests a property more problematic than interpretation-dependence. On this reading, in order to know the speaker's communicative intention in uttering an expression, one must antecedently know not just its linguistic meaning of this expression, but also its 'semantic value', i.e. its content in context. Other authors explicitly make this claim. In the following passage, Soames reflects on communicative intentions *qua* narrow contextual facts. In the first sentence, 'context' designates a sequence of parameters that includes the speaker's referent. In the second sentence, 'context' designates a sequence of parameters that includes only established narrow contextual facts:

> … to find out what context (in this sense of 'context') is relevant to understanding a speaker's remark, one often must *first* find out what the speaker is saying. As a result, one can't *explain* how speaker-hearers use their knowledge of meaning to interpret what is said by claiming that they apply their knowledge of semantics [...] to the contexts given to them in conversation (since *these* contexts don't come with labels specifying the contents of expressions). (Soames 2009, footnote 3 p.280, author's emphasis)

According to Soames, the fact that recognising the speaker's communicative intention in uttering an expression requires antecedently knowing the content of this expression clashes with a plausible view of the interpretation of indexicals. According to this view, we come to know the content of an indexical in context by antecedently knowing (i) the conventional meaning of this indexical and (ii) the contextual fact invoked by this meaning. It is not clear that there is an alternative to this view: given the kind of meaning that indexical expressions carry, it seems that things must go this way. But if communicative intentions play a narrow role for an expression *e*, and if recognising the speaker's communicative intention with *e*

[19] Qualifications would be needed here. It is not quite right that a person cannot know the speaker's communicative intentions without engaging *herself* in utterance-interpretation. As noted by Evans, there are *"extraneous or abnormal ways of gaining knowledge"* about an utterance (Evans 1982, p.314). Someone can do the interpretive work for me and then tell me what the speaker meant. I will then know what the speaker meant. However, *someone* must interpret the speaker's utterance in order for the communicative intentions of the speaker to be known. Secondly, the typical relation I describe between access to linguistic meaning and recognition of communicative intentions is typical to the extent that it holds for mature speakers of a language. By contrast, when it comes to learning a language it is plausible that the relation goes the other way around: one learns the meaning of words by recognising the communicative intentions of speakers who use these words. Some leading theories of language acquisition give pride of place to the recognition of communicative intentions (Bloom 2001, Tomasello 2010).

requires antecedently knowing the content of *e* in context, then the plausible view of the interpretation of indexicals cannot be right. We get a *reductio* of the thesis that communicative intentions play a narrow role.[20]

Gauker formulates the same worry differently: since knowing the content of an indexical in context requires antecedently knowing the contextual fact invoked by its meaning, and since knowing the communicative intention attached to the use of an expression requires antecedently knowing the content of this expression in context, interpreting an indexical endowed with an intentionalist meaning is an unfulfillable task. In Gauker's own words:

> My objection to this theory [communicative intentionalism] is that it renders the reference of demonstratives inaccessible to hearers. In order to identify the referent of a demonstrative the hearer will have to figure out what the speaker intended to refer to. But apart from an independent interpretation of the speaker's words, hearers will typically be in no position to do that. (Gauker 2008, p.363)

I find these arguments unconvincing. I am not convinced by the premise that recognising the speaker's communicative intention in uttering an expression requires antecedently knowing the content of this expression in context. But let me first identify a red herring with regard to this premise.

The quoted authors make a *conceptual* point about communicative intentions *qua* narrow contextual facts. But they assume a specific *formal* implementation of the idea of narrow contextual facts. This might lead us to think that the critical premise is motivated by matters of formal implementation, but this would be a mistake. On the formal implementation the quoted authors assume, the same entity - the speaker's referent - is both a parameter in the index *and* the value of the meaning-function associated with the relevant indexical expression. The meaning-function takes as an argument an index that contains the speaker's referent, and returns the speaker's referent as its value. Stokke observes that this formalism somewhat blurs the distinction between content-facts on the one hand, and the narrow contextual facts that determine content-facts on the other hand (Stokke 2010). However, the crucial thing to note here is that this issue does not afflict communicative intentions *specifically*. For example, on the standard formal implementation of the semantics for 'I', the speaker herself is a parameter, and the meaning-function associated with 'I' returns the speaker. Yet nobody claims that the interpreter must antecedently know that 'I' refers to *o* in context *c* in order to know that *o* is the speaker in *c*. The premise that knowing what the speaker communicatively intends with an expression requires antecedently knowing the expression's content is not motivated by the usual formal implementation of narrow contextual facts.

Why then do several authors claim that recognising the speaker's communicative intention in uttering an expression requires antecedently knowing the content of this expression in

[20] Recanati makes a very similar point in an earlier article written in French (Recanati 2001, pp.198-199).

context? Interpretation-dependence might be part of the motivation, but interpretation-dependence alone will not suffice. Indeed, it does not follow from the fact that communicative intentions are interpretation-dependent that recognising communicative intentions requires antecedent knowledge of content. To see this, it will help to consider an intentionalist Neale-Schiffer-style metasemantics for an expression e:

> *x* is the referent of expression *e* on an occasion of use *o* if and only if:
> (1) C*x*
> (2) the speaker refers to *x* with *e* on *o*

The interpreter who wants to ascertain condition (2) will use all sorts of facts to do so, and one of them will be the constraint C set by the meaning of expression *e*. If *e* is 'He', the interpreter will restrict her search for the speaker's referent to male beings. One aspect of the linguistic meaning of the expression - constraint C - is used to ascertain the contextual fact - speaker reference - which together with this aspect of meaning determines the content of the expression in context. But the content is not ascertained antecedently to the contextual fact that determines it. Neither interpretation-dependence nor dependence on a referential constraint - e.g. a restriction to male beings - makes the retrieval of the content of expressions with intentionalist meanings impossible. The worry that communicative intentionalism commits its proponents to an unfulfillable interpretive task seems to me misguided, as does the claim that recognising the speaker's communicative intention in uttering an expression requires antecedently knowing the content of this expression.

There is yet a further argument to be extracted from Brandom's quote. Brandom writes that the interpretation-dependence of communicative intentions entails that they cannot act as *inputs* to the operation of assigning 'semantic values' (i.e. contents) to expressions in context. Let me explain this. Inputs are pieces of information that are fed to a computational system. The system performs computations on these inputs using proprietary information, and perhaps using further information gleaned from other systems in the course of its operations. The system then returns its outputs. Now, suppose that there is a system in the human mind whose operations consist in accessing the linguistic meaning of uttered sub-sentential expressions, assigning contents to these expressions in the context of utterance on the basis of linguistic meaning, and composing these contents to return a content for the whole uttered sentence. Call this system the *semantic system*. If the meaning of an expression appeals to the communicative intention of the speaker, then information about this intention must become available to the interpreter when assigning a content to the relevant expression. This piece of information must be retrieved either before or in the course of the operations of the semantic system. In other words, information about the communicative intention is either an input to the semantic system, or it is gleaned from other systems in the course of the operations of the semantic system. A computational gloss on *interpretation-dependence* is that semantic operations must be antecedently initiated in order to recognise the speaker's communicative intention. If this is right, then communicative intentions cannot act as *inputs* to the semantic system. The remaining question is whether the second option - information about communicative intentions is gleaned from other systems in the course of semantic operations

- is plausible. If neither option is plausible, then communicative intentions cannot act as narrow contextual facts.

Perhaps Brandom assumes that the second option is not plausible, and thus focuses on making the point that the first option - communicative intentions *qua* inputs to the semantic system - is not plausible. In any case, for this argument against communicative intentions *qua* narrow contextual facts to go through, one needs to establish the premise that information about communicative intentions cannot be gleaned from other systems by the semantic system in the course of its operations. I turn to this premise in the next section.

3. The insulation of semantics from pragmatics

Let me restate a widely accepted view of what the semantic system does. Once sounds have been recognised as phonemes and words, and once a sequence of words has been recognised as a grammatical sentence and been assigned a syntactic tree, this syntactic tree is sent as an input to the semantic system, perhaps along with additional information in the form of parameters. The semantic system finds in its proprietary database the meaning of the smallest constituents on the syntactic tree. It then assigns *denotations* - or *compositional values* - to these constituents in accordance with their linguistic meaning. I am using 'denotation' and 'compositional value' interchangeably to designate the entity that enters composition, assuming a compositional semantics. The formulation 'in accordance with linguistic meaning' leaves it open whether the compositional value of an expression is identical with the meaning of the expression-type, or with something determined by it - such as its content in context.[21] The semantic system then combines sub-sentential denotations according to compositional rules, until a denotation is assigned to the whole sentence.

Let us call *pragmatic reasoning* the cognitive task of interpreting a speaker's communicative intentions, however this task is carried out. What the argument sketched at the end of the previous section requires is that the operations of the semantic system - which we may call *semantic decoding* - be insulated from pragmatic reasoning. From taking its input - say the syntactic tree of the uttered sentence plus several parameters - to returning its final output - the denotation of the uttered sentence - the system that underlies semantic decoding has no contact with the systems that underlie pragmatic reasoning. Assuming the view of semantic decoding presented in the previous paragraph, this translates into the thesis that no pragmatic reasoning intervenes either in the course of assigning denotations to sub-sentential constituents or in the course of composing these denotations.

The view that semantic decoding is insulated from pragmatic reasoning should be distinguished from the weaker view that the calculation of a sentence-denotation is free from

[21] In Kaplan's theory for example, meanings of expression-types ('characters') do not compose: contents in context do (Kaplan 1989a). Lewis suggests a semantics in which meanings of expression-types compose (Lewis 1980).

any pragmatic reasoning that is not *required to assign a denotation to some sentential constituent in accordance with linguistic meaning*. On this weaker view, pragmatic reasoning intervenes in-composition if and only if the assignment of a compositional value to a constituent demands it. Recanati attributes the weaker view to Stanley (Recanati 2010, chapter 1), and Stanley accepts this attribution (Stanley 2005). The weaker view is consistent with an intentionalist metasemantics for some expressions, assuming that this metasemantics is encoded as the linguistic meaning of the expressions.

As assumed in this chapter, communicative intentionalism about linguistic expression *e* has it that the meaning of *e* instructs interpreters to identify the speaker's communicative intention in uttering *e* in order to assign a content to *e* in a context of utterance. This means that information about the speaker's communicative intention in uttering *e* must be available when assigning a content to *e* in context. However, communicative intentionalism is not strictly inconsistent with the insulation of semantics from pragmatis. Indeed, if the compositional value of sub-sentential expressions is their linguistic meaning rather than their content in context, then content is assigned *after the semantic system has returned its final output.* Given an intentionalist meaning, information about the communicative intention of the speaker is required to assign a content to an expression with intentionalist meaning, but this piece of information is retrieved *post-composition*, which is consistent with the insulation thesis. In summary, communicative intentionalism is inconsistent with the three following theses:

(1) The compositional value of linguistic expressions is their content in context.
(2) Communicative intentions are not inputs to the semantic system.
(3) Semantic decoding is insulated from pragmatic reasoning (the insulation thesis).[22]

Now, why should one endorse the insulation thesis? Some authors have appealed to the *modularity* of the semantic system (Borg 2004, Harris 2020). An overall argument from modularity against communicative intentionalism runs as follows:

Premise 1: The semantic system is modular.
Premise 2: If the semantic system is modular, then it is insulated from pragmatic reasoning.
Conclusion 1: The semantic system is insulated from pragmatic reasoning.

Premise 3: If the semantic system is insulated from pragmatic reasoning and theses (1)-(2) above are true, then there cannot be reference-determining communicative intentions.
Premise 4: Theses (1)-(2) above are true.
Conclusion 2: There cannot be reference-determining communicative intentions.[23]

[22] Robbins thinks that the truth of the insulation thesis does not threaten communicative intentionalism because communicative intentions can be regarded as *inputs* to the semantic system (2007, pp.312-313). As I explained in the previous section, the interpretation-dependence of communicative intentions makes them unlikely inputs. Harris accepts (2)-(3) and some form of communicative intentionalism, and thus rejects (1). He goes on to sketch an alternative compositional semantics in which something other than content composes (Harris 2020).
[23] The argument up until Conclusion 1 is that of the second chapter of (Borg 2004).

I want to focus on the first part of the overall argument, which is an argument *for insulation from modularity*. I need to say a few words about modularity first. The idea underlying the notion of modularity is that of specialised and broadly autonomous computational systems in the human mind (Fodor 1983). Fodor associates the notion of modularity with a list of system-properties, but according to him only some of the properties on this list are necessary for a system to count as modular. One way to understand this view is that some of the properties on Fodor's list are possessed by *all* specialised systems in the human mind, while some other properties are variably instantiated. To make matters more complicated, there is disagreement about which properties in Fodor's list are essential to modularity and in what form (Carruthers 2006, Coltheart 1999). This disagreement may be understood as one about what the autonomy of a specialised system consists in exactly.

The only property on Fodor's list that is directly relevant to insulation is *encapsulation*.[24] There are absolute and relative notions of encapsulation. According to the absolute notion, an encapsulated system is one that does not use information from any other system in the course of its operations. An encapsulated system only uses its inputs, its proprietary database, and its own processing resources to return its outputs. As for the relative notion of encapsulation, a system A is encapsulated from a system B if A does not use information from B in the course of its operations.[25] One form of relative encapsulation which looms large in the literature, and which might be Fodor's notion, is encapsulation from central cognition - also called 'cognitive impenetrability' (Coltheart 1999). Central cognition is that part of the human mind where conceptual representations, including complex ones such as propositional attitudes, are consciously entertained and manipulated. Fodor regards encapsulation as an essential property of specialised systems, but both the absolute encapsulation and the relative encapsulation from central cognition of specialised systems have been put into question (Coltheart 1999, Robbins 2017).

The relevance of encapsulation to the insulation thesis is clear. The insulation thesis is just the thesis that the semantic system is encapsulated from whatever systems underlie pragmatic reasoning - henceforth *pragmatic systems*. The absolute encapsulation of the semantic system thus entails the insulation thesis. As for the relative encapsulation of the semantic system *from central cognition*, things are not quite as straightforward. If central cognition is in charge of pragmatic reasoning, the relative encapsulation of the semantic system from central cognition entails insulation. But if one or more specialised systems are in charge of pragmatic reasoning, the encapsulation of the semantic system from central cognition is consistent with some pragmatic reasoning taking place in the course of semantic decoding. The place of

[24] I will not argue for the exclusivity aspect of this claim, as it would require going through Fodor's entire list. Let me just flag here that the *inaccessibility* of the semantic system could have interesting consequences on the intervention of pragmatic reasoning in semantic operations. As with encapsulation, there are absolute and relative notions of inaccessibility. In addition, there are more or less demanding notions of inaccessibility of a system depending on what exactly other systems do not have access to (Carruthers 2006, Robbins 2017).

[25] According to Carruthers, another distinction can be made between narrow-scope and wide-scope encapsulation (Carruthers 2006, pp.57-59).

pragmatic reasoning in the architecture of the human mind is a vexed question which I will not tackle here.[26]

With this background in place, it seems to me clear that arguing from the modularity of the semantic system to the insulation thesis is not an optimal strategy. If only encapsulation among properties associated with modularity is directly relevant to insulation, one should simply focus on establishing that the semantic system is encapsulated to a degree that entails its insulation from pragmatic reasoning. Both Borg and Harris make it clear that they use 'modularity' in the Fodorian sense, and hence that they take encapsulation to be essential to modularity (Borg 2004, Harris 2020). However, these authors spend time arguing that the semantic system has all the other properties on Fodor's list. This seems to me a distraction. The fact that the semantic system instantiates properties P1, P2, P3… other than encapsulation on Fodor's list does not tell us anything about encapsulation unless *any system that instantiates P1, P2, P3… is encapsulated*.[27] Regardless of Fodor's view on the pattern of co-instantiation of properties on his list, the italicised claim is a substantial one that needs to be argued for. Without such an argument, the instantiation of P1, P2, P3… by a system makes it at best more likely that the system is encapsulated than if the system did not instantiate these properties. In conclusion, one should not argue for insulation from modularity, or from modularity-associated properties other than encapsulation.

Even the relation between encapsulation and insulation is not straightforward. As we have just seen, insulation from pragmatic reasoning follows from the encapsulation of the semantic system *from central cognition* only if central cognition is in charge of pragmatic reasoning. Only *absolute* encapsulation immediately entails insulation. However, this does not mean that arguing for insulation from absolute encapsulation is a good idea. The absolute encapsulation of a system S is a universal claim: for any system x distinct from S, S does not use information from x. As for the insulation thesis, it is just the thesis that the semantic system is encapsulated from pragmatic systems. It seems absurd to try and establish the insulation thesis by establishing first that the semantic system is encapsulated from a range of systems other than pragmatic systems, when one could instead directly establish that the semantic system is encapsulated from pragmatic systems. Hence, both modularity and encapsulation are red herrings when it comes to arguing for the insulation of semantic decoding from pragmatic reasoning. If one wants to argue for the insulation thesis, one should do so directly.

---

[26] On the one hand, it seems that pragmatic reasoning has no limit on the information that it can use, and that it is abductive (rather than algorithmic). This constitutes the classical case for pragmatic reasoning as a central process. On the other hand, there is a trend for conceiving of most pragmatic reasoning as realised by fast, heuristic-dominated processes which can be handled by one or more specialised systems (Sperber & Wilson 2002).

[27] Carruthers argues against this claim (Carruthers 2006). According to him, computational systems that instantiate other properties on Fodor's list are *frugal* systems, but frugality does not require encapsulation (whether absolute encapsulation or encapsulation from central cognition). It is therefore very likely that there are systems in the human mind that do not instantiate encapsulation while instantiating other properties on Fodor's list.

There seem to be two possible sources of evidence relevant to the insulation thesis. Whether or not pragmatic reasoning is needed to assign a content to context-sensitive expressions, it is relatively uncontroversial that pragmatic reasoning is sometimes needed to assign a denotation to a sentence. For example, pragmatic reasoning is sometimes needed to resolve lexical and syntactic ambiguities in a sentence.[28] When pragmatic reasoning is needed, disambiguation demands attention to the overall meaning of the sentence at hand. For example, arriving at a reading of 'Chestnuts are shedding their leaves' in which 'chestnuts' stands for the trees requires knowledge about the kind of thing that sheds leaves.[29]

Not all views of the cognitive processes underlying disambiguation are consistent with the insulation thesis. The insulation thesis allows that the semantic system composes every possible sentence-denotation and then delivers them all to the pragmatic system, which selects one sentence-denotation. Let us call this view the 'naive' serial view of disambiguation. The insulation thesis also allows that the semantic system selects one possible denotation on the basis of proprietary information and proprietary heuristics (for example, information about the relative frequency of competing meanings of a word and a general heuristic that selects the 'dominant' meaning) before delivering a single sentence-denotation to the pragmatic system. On this second view, the pragmatic system checks the unique denotation delivered by the semantic system and sends it back only if it is contextually implausible. For both possible disambiguation processes, the pragmatic system contributes to disambiguation *after composition*; that is, after the semantic system delivers its final output. The insulation thesis does not allow such a contribution to take place before denotations for the whole sentence are returned. Hence, a third view of disambiguation according to which a denotation for an ambiguous expression is selected in view of the denotation of other expressions in the sentence *in the midst of composition* is not consistent with the insulation thesis.

Empirical studies recording brain activity in ambiguity processing seem to favour some theories of ambiguity processing over others. When it comes to homonymy with two candidate meanings, it seems that the two candidate meanings are initially accessed by the semantic system (Swaab et al. 2003). This means that lexical access is not influenced by considerations of contextual plausibility, although this does not preclude that one meaning is selected at some point before the semantic system returns a denotation for the whole sentence. The same study suggests that the contextually implausible meaning is quickly deactivated when it is the less frequent meaning (the 'subordinate' one), while it stays partly activated when it is the more frequent meaning (the 'dominant' one). In other words, it looks like the dominant meaning is given the benefit of the doubt to an extent that the subordinate meaning is not. These results seem to be consistent with the idea that the meaning deemed dominant by the semantic system produces the denotation that enters composition, while the subordinate meaning is given a proper run only if the dominant meaning yields an implausible

[28] I write 'sometimes' because, as noted by Borg, some ambiguities can be resolved by applying only grammatical knowledge or proprietary selection heuristics (Borg 2004, p.143). In addition, disambiguation is sometimes guided by both selection heuristics and pragmatic reasoning (Swaab et al. 2003).

[29] This example is taken from (Geurts & Rubio-Fernandez 2015).

compositional result. Another study indicates that areas of the brain previously discovered to be involved in lexical disambiguation are more intensely activated when the contextually plausible meaning is the subordinate one than when it is the dominant one (Zempleni et al. 2007). Strikingly, there is almost no difference in the activation of these areas between (i) the processing of a sentence containing a homonym whose dominant meaning is the contextually plausible one and (ii) the processing of a sentence containing no ambiguous word. This pattern is easily explained if the dominant meaning is selected early on and disambiguation is a correction process that takes place only when the dominant meaning yields a contextually implausible compositional result. This pattern of evidence puts pressure on the 'naive' serial view according to which all possible denotations for an ambiguous word enter composition regardless of frequency and contextual plausibility.

The study by Zempleni and co-authors therefore favours the view that the semantic system contributes to lexical disambiguation by selecting one meaning on probabilistic grounds, while the pragmatic system contributes by assessing the contextual plausibility of the compositional result of this selection. This view might seem to favour the insulation thesis, but it is actually consistent with its negation. The assessment of contextual plausibility by the pragmatic system could be carried out either during composition - in the midst of the operations of the semantic system - or post-composition - after the semantic system has delivered its final output. For example, with a sentence such as 'The chestnuts are shedding their leaves on a cold November morning', checking contextual plausibility could take place as soon as the denotation of 'The chestnuts are shedding their leaves' is calculated, rather than after a denotation for the whole sentence is calculated. This makes me sceptical of the possibility of finding processing evidence for or against the insulation thesis. Processing evidence can discriminate between different theories of the processes involving some pragmatic reasoning, but it is not clear that it can discriminate between different views of where exactly pragmatic reasoning intervenes in these processes, i.e. in-composition or post-composition.

There is another source of evidence which could be relevant to the insulation thesis. One finds on Fodor's list of system-properties associated with modularity the property of *functional dissociability*. Here again there may be absolute and relative variants of the notion, but I will focus on the relative one. A system S1 is functionally dissociable from another system S2 if S1 can be impaired without S2 being impaired. The functional dissociability of S1 from S2 may be thought to constitute evidence for the encapsulation of S2 from S1. Indeed, if S2 uses information from S1 to carry out its operations, S1's impairment should hinder S2's operations. So if S1's impairment is not accompanied by an impairment of S2, it is likely S2 does not recruit S1 in its operations. In the context of this section, the functional dissociability of pragmatic competence from semantic competence may constitute evidence that the system underlying semantic competence is encapsulated from the systems underlying pragmatic competence.

Prima facie evidence of the dissociability of pragmatic competence from semantic competence can be found in people with autism disorder (Frith & Happé 1994), people with

schizophrenia (Langdon et al. 2002), and people with specific brain lesions (Happé et al., 1999). Focusing on people with autism disorder, it is widely thought that their capacity to represent the mental states of others is not as developed as that of neurotypical people (Baron-Cohen 1995). This capacity is known under different names in the literature: 'mind-reading', 'theory of mind', 'mentalising', and 'metarepresentation'. On the face of it, the pragmatic capacity is a capacity to attribute intentions of a certain kind to other people, and so it overlaps with the 'mentalising' capacity. One should then expect some degree of pragmatic impairment in people with autism. This is borne out by empirical evidence. People with autism struggle with the production and comprehension of various forms of speech that involves going beyond literal meaning, including irony, metaphors, jokes, and conversational implicatures (Happé 1993). Crucially for our present purposes, this difficulty is found in autistic individuals whose linguistic performance, including semantic performance, matches that of their neurotypical peers. Despite some distinctive patterns, these individuals produce and understand literal speech as smoothly as neurotypical individuals. This part of the autism data set seems to support the insulation thesis. It looks like semantic competence can be left roughly intact while pragmatic competence is substantially impaired. This suggests that the semantic system is encapsulated from the pragmatic system.

Other findings paint a more confusing picture. Some studies indicate that linguistically proficient autistic people perform well on the production and comprehension of non-figurative speech which plausibly requires some pragmatic reasoning (de Villiers et al 2007, Kissine 2012). This includes the production and comprehension of ambiguous sentences in context (Norbury 2005). A possible response to this part of the data set is to deny that the processes underlying disambiguation typically involve representing the communicative intentions of speakers - or any metarepresentation for that matter. After all, one can disambiguate e.g. 'A doctor lives in every city' without thinking about the communicative intention of its utterer. Only 'world-knowledge' is necessary, in this case the arity of the *living in* relation (Saba and Corriveau 2001). This response is acceptable, but one should keep in mind that comprehension phenomena such as disambiguation involve *non-linguistic* processes: information must be gleaned from outside the semantic and syntactic systems in order to arrive at a truth-conditional interpretation for the target sentence, whether that information is about the mental states of the speaker or mere 'world-knowledge'.

With this qualification in mind, the following interpretation of the whole data set seems to me plausible. Although the aim of the non-linguistic part of utterance-interpretation is to recognise the speaker's communicative intentions, this part is carried out in places by *mentalising-free processes* (Kissine 2012). Not only autistic people, but also their neurotypical peers fulfil some pragmatic tasks without attending to the speaker's mental states. For example, the *"automatic application of a relevance-based procedure"* might be good enough in many cases where linguistic information gives out (Sperber & Wilson 2002). The absence of a mentalising burden for the interpreter in some pragmatic tasks makes sense if the speaker carries this burden. Indeed, if the speaker tailors her utterance to the interpreter's perspective, the interpreter need not herself consider the speaker's perspective for

communication to succeed.[30] The hypothesis concerning autistic people is then that among pragmatic tasks they struggle only with those tasks for which consideration of the speaker's mental states is ineliminable (Kissine 2012). In addition, the view according to which the mentalising burden weighs mostly on the speaker predicts that one should find an asymmetry between correct production and correct comprehension of some constructions requiring pragmatic reasoning in people with autism. There is some evidence of this asymmetry. Children with autism tend to use third-person pronouns in contexts where their reference is difficult to ascertain for their audience, but avoid them in places where their reference would be easily ascertained (Novogrodsky & Edelson 2016). On the other hand, it seems that children with autism have no comprehension problem with third-person pronouns (Nagano et al, 2021). This is what we should expect if the mentalising burden is taken on by the speaker when it comes to third-person pronouns.[31]

The case of autism thus provides some evidence for the functional dissociability mentalising-heavy, non-linguistic interpretive processes from semantics. This is evidence that the semantic system is encapsulated from whatever system is in charge of mentalising-heavy processes. This is in turn evidence for qualified versions of the insulation thesis, according to which no mentalising-heavy pragmatics intervene before the semantic system returns a denotation for a whole sentence. If postulating intentionalist meanings for sub-sentential expressions commits one to heavy mentalising on the part of interpreters in the assignment of contents to these expressions, then available evidence suggests that one refrains from postulating intentionalist meanings (assuming that the semantic system composes *contents*). However, it is not clear to me that an intentionalist meaning for e.g. 'That' or 'He' demands heavy mentalising from interpreters. In conclusion, it seems to me that there is no decisive evidence of architectural features of the human mind that are inconsistent with communicative intentionalism.

4. Redundant meanings

Several authors have suggested that linguistic meanings appealing to the communicative intentions of speakers are somehow *redundant*. We will shortly see what this charge amounts to, but let me start with a claim made by Bach. The intention constitutive of speaker reference *"is part of a speaker's total communicative intention, and that makes it pragmatic in import, not semantic"* (Bach 2017, p.58). As I understand Bach, the fact that communicative intentions have *pragmatic import* has the following consequence. Lists of conditions that include speaker reference and are presented as the determinants of a property of expressions track a property of speakers. Take a Neale-Schiffer-style metasemantics:

[30] Of course, some expressions have a meaning such that interpreting utterances of them essentially requires taking the speaker's perspective into account. 'Here', 'I', and 'You' are good examples.

[31] This simple asymmetry picture will not apply to context-sensitive expressions whose very meaning is linked to the perspective of the speaker (see previous footnote). Difficulty in representing the perspective of others should cause problems in mastering these expressions, and thus cause downstream problems with both production and comprehension.

Expression *e* refers to object *x* on occasion of use *o* if and only if:
(1) C*x* (where C is the constraint set by the conventional meaning of *e*)
(2) The speaker refers to *x* with *e* on *o*

This list of conditions tracks the pragmatic property of a speaker *literally referring* to an object:

Speaker S literally refers to *x* with *e* on occasion of use *o* if and only if:
(1) C*x* (where C is the constraint set by the conventional meaning of *e*)
(2) the speaker refers to *x* with *e* on *o*

In the same vein, Bach argues that Stokke's proposed metasemantics for demonstratives tracks a pragmatic property (Bach 2017, pp.62-65; Stokke 2010). Here is Stokke's metasemantics:

Expression *e* refers to object *x* on occasion of use *o* if and only if:
(1) An attentive and competent audience would take the speaker to have referred to *x* with *e* on *o*
(2) The speaker refers to *x* with *e* on o

These conditions track the pragmatic property of (something like) *skillfully referring*:

Speaker S skilfully refers to *x* with *e* on occasion of use *o* if and only if:
(1) An attentive and competent audience would take the speaker to have referred to *x* with *e* on *o*
(2) the speaker refers to *x* with *e* on *o*

Now, Bach does not argue that properties of speakers could not in principle do double-duty as properties of expressions. However, to do double-duty as properties of expressions, these properties must have their determinants encoded in the linguistic meaning of the relevant expressions. To be semantic in the sense of being properties of expressions, these properties must be semantic in the sense of being linguistically determined. Bach therefore challenges intentionalists to justify the postulation of linguistic meanings that determine a property which can independently be construed as *not* determined by linguistic meaning.

This challenge does not exactly amount to a claim of redundancy, but it is clear that Bach regards meanings appealing to speaker reference as *theoretically redundant*. A linguistic meaning for an expression is theoretically redundant just when its postulation brings no explanatory gain. Potential *explananda* of a linguistic meaning include: intuitions about the truth-conditions of sentences containing the relevant expression, intuitions about entailment patterns between sentences, communication with the expression… Of course, linguistic meanings might not carry the whole explanatory burden on their own. For example, general principles governing talk exchanges might play a role with regard to the aforementioned

explananda. Grice famously argued that taking into account these principles makes some putative linguistic meanings theoretically redundant (Grice 1989).

There is a more specific sense in which a linguistic meaning can be redundant. A postulated meaning for an expression can be *communicatively* redundant: without it, speakers of the relevant language would use and interpret uses of this expression as they actually do. The communicative redundancy of a meaning may be seen as a restricted form of theoretical redundancy focusing on one *explanandum*: how speakers use and interpret uses of the expression as they do. Communicative redundancy plays a role in several arguments in the literature. Bach thinks that any meaning for demonstratives that associates a demonstrative with an object in a context of utterance is communicatively redundant. A connection between communicative redundancy and overall theoretical redundancy is made in the following passage:

> You can still attribute referential semantic values to these expressions if you like—say, for formal semantic purposes; but they play no role in the process whereby speakers make references and hearers identify what speakers refer to when they use these expressions (Bach 2017, p.66)

Nowak also argues that such meanings for demonstratives are theoretically redundant because they are communicatively redundant (Nowak 2019). Nowak calls the view that demonstratives have a meaning that associates the expression with an object in a context of utterance the *"lexical contextualist view"*. Here is Nowak's argument:

> In a nutshell, the problem is that on the lexical contextualist view, instructions for finding the referent of a demonstrative are encoded in its meaning. But those instructions are instructions no reasonable interpreter would need: the relevant constraints fall naturally out of the structure of the interpretive task people are faced with when they encounter demonstratives. So, lexicalizing those constraints amounts to unnecessary double-counting. (Nowak 2019, p.197)

If we call *English'* a language that includes such meanings:

> The constraints that are lexicalized in English' are constraints that we get for free by basic pragmatic reasoning. By encoding those constraints in the semantics, we exchange a simpler theory for a more complicated one that gains us no explanatory power.

Here I will treat these arguments as arguments against *intentionalist* meanings that associate expression with object in context specifically. What is wrong with these meanings then? Well, since a non-linguistic *pragmatic* capacity springs into action whenever behaviour is recognised as communicative behaviour, interpreters would interpret demonstratives as they actually do if there was no linguistic meaning instructing them to recognise the speaker's communicative intention. Interpreters are able to recognise communicative behaviour when

they see it, and know that the task at hand when faced with communicative behaviour is to recognise the speaker's communicative intention. Hence, when faced with a linguistic communicative act, interpreters set out to infer the utterer's communicative intention in uttering linguistic expressions. They do not need the meaning of an expression to trigger this process, since the mere recognition that the expression has been uttered independently triggers it. Therefore, meanings appealing to communicative intentions are communicatively redundant. Nowak further concludes that this communicative redundancy makes intentionalist meanings theoretically redundant.

I am convinced by the claim that intentionalist meanings are communicatively redundant. But contrary to what Nowak suggests, communicative redundancy does not entail theoretical redundancy. There are other potential *explananda* for a putative linguistic meaning than how content can be communicated. For example, pre-theoretical judgements about the truth-conditions of sentences are a core *explanandum* of semantics in the truth-conditional tradition. Hence, Nowak does not acknowledge that he has set up a conflict between two desiderata for linguistic meanings: communicative non-redundancy and explaining truth-condition judgements.[32] Nowak further holds that a meaning for demonstratives is communicatively *non*-redundant only if it does not determine an object in context. This means that two theoretical desiderata - (i) being communicatively non-redundant and (ii) explaining truth-condition judgements - cannot be jointly satisfied by a meaning for demonstratives. Nowak's conclusion of theoretical redundancy is thus conditional on rejecting (ii) as a desideratum. Bach is more transparent on this score. He acknowledges that a communicatively non-redundant meaning for demonstratives yields a non-truth-conditional denotation for sentences containing demonstratives, but he explicitly rejects (ii) as a desideratum (Bach 2017, p.67).

One avenue for the foe of communicative intentionalism is therefore to argue against desideratum (ii) on independent grounds. Bach and other authors have gone down this path (Bach 1994b, Carston 2002, Recanati 2004, Sperber & Wilson 1995). Another argumentative path can be taken at this juncture though. One could make an independent case for communicative non-redundancy as a theoretical desideratum for linguistic meanings. This would be an argument for the claim that there cannot be communicatively redundant meanings (where the strength of 'cannot' should be specified).

Before pursuing this line of thought, one can pressure intentionalists to justify the postulation of communicatively redundant meanings. As Nowak makes it clear, intentionalist meanings trigger an operation - the recognition of the speaker's communicative intention - that is independently triggered by the recognition of the speaker's behaviour as communicative behaviour. Hence, under a regime of communicative intentionalism the process of utterance-interpretation involves a *redundant command*. A mental process includes a redundant command when one and the same operation is independently requested by several

[32] We saw in the previous chapter that intentionalism makes predictions that clash with pre-theoretical judgements in a number of cases. So to be fair to Nowak it is not clear where intentionalism could earn its theoretical keep. I put aside intentionalism's predictive problem in this chapter.

components of this process. Unless redundant commands are pervasive in human cognition, intentionalists owe us an explanation of why utterance-interpretation involves one. In addition, intentionalists owe us a more specific explanation of why the semantic component of utterance-interpretation should issue a command already issued by the pragmatic component, when it does not issue commands already issued by other components. For example, no theorist suggests that the meaning of some expression makes a request to the *phonological* component of utterance-interpretation to distinguish phonemes in the noises made by the speaker. And no theorist suggests that the meaning of some expression demands that the *grammatical* component of utterance-interpretation place the expression on a syntactic tree. Why should things be different when it comes to the *pragmatic* component of utterance-interpretation?

The next step is to argue directly against the idea that the semantic component of utterance-interpretation issues commands already issued by the pragmatic component. Bach argues that given a certain view of the linguistic meaning of referential expressions, these expressions could not have intentionalist meanings (Bach 1994a, pp.178-179). According to Bach, the meaning of a referential expression restricts speaker reference to objects of a certain class. The meaning of 'He' restricts speaker reference to male beings. The meaning of 'I' restricts speaker reference to the speaker. This view of meaning as a constraint imposed on the communicative intentions of those who use them can be captured with *rules of use*. For example, the linguistic meaning of 'I' and 'He' can be captured by the following rules: use 'I' only when you refer to yourself; use 'He' only when you refer to a male being.[33] If this picture is endorsed, the problem with intentionalist meanings is simple. Such meanings restrict speaker reference to what the speaker refers to, which is just to say that they do not restrict speaker reference. Given a pure intentionalist metasemantics for an expression *e*, the rule of use for *e* is: use *e* to refer to what you refer to. Given an impure intentionalist metasemantics à la Neale-Schiffer - with other conditions than a speaker reference condition - the rule has a useless component: use *e* to refer to something which is C *and which you refer to*.

The question raised by Bach's argument is why we should accept his view of the linguistic meaning of referential expressions as mere restrictions on speaker reference. Answering this question is the key to the conclusion that there cannot be communicatively redundant meanings. Nowak claims that referential expressions cannot encode more than restrictions on speaker reference given how language is acquired by children - first qualification - according to some theories of language acquisition - second qualification (Nowak 2019, pp.200-203). The first qualification specifies the strength of 'cannot' in the claim that there cannot be communicatively redundant meanings of the sort posited by intentionalists. As for the second

[33] The implementation of this idea in model-theoretic semantics is controversial. Assuming that these restriction-meanings compose, one option is to model different restriction-meanings as different restrictions of the function from variable assignments to objects (Nowak 2019, p.205). Some authors doubt that entities such as variable assignments are represented in our semantic system (Jacobson 1999), and some authors doubt that central cognition could handle outputs of semantic composition that traffic in variable assignments (Harris 2020). Harris puts forward an alternative formal representation of restriction-meanings.

qualification, Nowak uses (Gillette et al. 1999) as a friendly theory of language acquisition. One could also mention (Bloom 2001) and (Tomasello 2010). According to these theories, the capacity to recognise speaker reference - or some embryonic form of this capacity - plays a significant role in language acquisition. Children learn the meaning of words partly by recognising what object the speaker is drawing their attention to in the situation of utterance. If this is right, it is hard to see how children could end up associating some particular expression with the instruction to find out what the speaker is drawing their attention to. Furthermore, when children come to understand that an expression is a referential expression - i.e. one whose very role it is to draw attention to a particular object - one possible gloss on this acquisition-step is that they associate the expression with semantic type *e* - the type of objects. Once this association is made, a further association with an instruction to find out what object the speaker is referring to is not necessary, and therefore this further association is just not made. This does not mean that other associations are not made: to master the third-person pronoun 'It' for instance, one must come to understand that 'It' is used to refer to inanimate objects.

This argument from language acquisition might have an evolutionary analogue, although what follows is very tentative. It is now widely accepted that great apes make gestures with communicative intentions, although ones apparently limited to imperatives and not including the more specific communicative intentions constitutive of speaker reference (Byrne et al. 2017, Tomasello 2019). Some embryonic form of the capacity to act with communicative intentions and to recognise such intentions was thus present in our evolutionary history before language as we know it emerged. Assuming that there was an intermediary stage in our evolutionary history, after divergence from other great apes, where gestures or gazes were used with intentions characteristic of speaker reference, and that these actions were interpreted as carrying such intentions, the development of human language piggybacked on this capacity for speaker reference and its recognition (Fitch 2010). Linguistic expressions used for referring thus emerged against a background of non-linguistic utterances involving speaker reference. To get a referring linguistic expression up and running, it was then only necessary for the expression to encode its referential function: its semantic type *e*. This encoding was enough for the utterance of the expression to trigger the search for the speaker's referent. Further encoding instructions to search for the speaker's referent were not necessary to get referential expressions up and running. It is then implausible that expressions encoding such instructions emerged as human language emerged.

More detailed arguments from language acquisition and language evolution should be offered, but it seems to me that communicative non-redundancy is a plausible constraint on linguistic meaning. Since linguistic meanings demanding the recognition of communicative intentions are communicatively redundant, such meanings are implausible.

Conclusion

The pre-theoretical divergence between what one says and what one means does not on its own undermine communicative intentionalism (section 1). The interpretation-dependence of communicative intentions does not count against communicative intentionalism, unless the semantic system is insulated from pragmatic systems (section 2). While there is some evidence that the semantic system is insulated from whatever system is in charge of mentalising-heavy pragmatic processes, it is not clear that communicative intentionalism is threatened by this insulation (section 3). However, linguistic meanings that invoke communicative intentions are communicatively redundant, and there are reasons to think that there are no communicatively redundant meanings in human languages (section 4).

Chapter 3: Conceptual problems for semantic intentionalism

Some intentionalists do not favour *communicative* intentions as determinants of semantic reference (King 2014; Speaks 2016, Viebahn 2020). The perceived need to do justice to the pre-theoretical distinction between what the speaker means and what the speaker says plays a role in this dislike (Viebahn 2020, pp.95-97). Upstream scepticism about Gricean communicative intentions might also play a role. In any case, some authors think that intentions other than the intentions constitutive of speaker reference determine the reference of demonstratives.

One could use the label 'semantic intentions' to designate the non-communicative intentions which determine the reference of demonstratives. The label would then not specify anything about the content of reference-determining intentions, except that they are not communicative. But communicative-intentions-averse intentionalists have something more specific in mind. The intentions they favour as determinants of reference, and which they call 'semantic', are intentions *about the properties of expressions*.[34] In the context of a debate about the facts that determine reference, the relevant property is the reference of an expression in context. Some intentionalists thus propose that the reference of expressions is determined by an intention about the reference of expressions. We can call this view *semantic intentionalism*.

I will not address every possible objection to semantic intentionalism here. Let me flag some important issues which will not be discussed in detail. First, some of the considerations against reference-determining communicative intentions discussed in the previous chapter might also threaten reference-determining semantic intentions. Here I ignore these potential overlaps. Secondly, just as the necessity of communicative intentions for speaker reference is controversial given the sophistication of communicative intentions, the necessity of semantic intentions for semantic reference is controversial given the sophistication of semantic intentions (Devitt 2022, Speaks 2016, Viebahn 2020). As I see it, over-intellectualisation worries crop up in philosophy whenever some phenomenon explainable in terms of contentful mental states is explained in terms of the familiar explicit propositional attitudes, for lack of a distinct vocabulary for non-explicit, non-conscious, contentful mental states. In other words, I think that the over-intellectualisation problem is a general problem of philosophical methodology rather than a problem for theories of the determination of reference. This is why I will ignore it here. Finally, semantic intentionalism yields conditional

[34] I use 'semantic intentions' differently from Viebahn, who uses it to designate whatever intentions determine the semantic content of expressions in context, be they audience-directed communicative intentions, intentions about the content of expressions, or something else (Viebahn 2020, p.94).

statements in which the same notion of reference appears on both the left-hand side and the right-hand side - e.g. *d* refers to *o* in *c* only if the speaker intends that *d* refers to *o* in *c*. This raises a circularity worry (Gauker 2008), which to my mind has been answered satisfactorily (Lewerentz and Marschall 2018).

In this chapter I focus on two problems for semantic intentionalism. The first problem is that the content usually given to semantic intentions is not fit to be the content of an intention. One must put forward a content for these intentions that is consistent with a plausible general theory of intention-content. The second problem is one of ad hocness: one cannot posit intentions about the reference of expressions just because one wants to be an intentionalist on the one hand, but does not want communicative intentions to do the reference-determining job on the other hand. One must provide independent reasons for thinking that speakers have intentions about the reference of the expressions they utter.

1. What is the content of intentions about reference?

King (2013, 2014) and Speaks (2016) characterise reference-determining semantic intentions as follows:

> (SI) The speaker intends object *o* to be the semantic value [i.e. reference] of demonstrative *d* in context *c*.

There seems to be a consensus around this formulation (Bach 2017, Lewerentz & Marschall 2018). This is surprising, to the extent that the content of (SI) does not seem appropriate as the content of an intention. Pre-theoretically, one intends *to do something*. Yet no action figures in the content of (SI). Intuitively, one cannot intend e.g. the temperature to be above 20 degrees, although one can intend to turn up the thermostat *so that the temperature is above 20 degrees*.

Beyond pre-theoretical judgements, most theories of intention-content would not tolerate (SI). The most popular theory holds that the content of intentions is a proposition that contains an agent S and an action-type *φ*. The content of intentions is then best represented as follows: S intends *that S φ's* (Campbell 2019). The main competitor of the propositional theory holds that the content of intentions is just the action-type *φ*, and is thus non-propositional (Baier 1970). This second theory follows the surface grammar of intention-attributions more closely: we report people as intending *to do something* (Campbell 2019).[35] Even beyond the two main theories of intention-content, it is hard to find a view that puts neither the intended act nor its agent in the content of intentions. Castaneda takes intention-contents to be 'prescriptions': these are *"the fundamental units of practical thinking, just as propositions are the*

[35] There is a further debate concerning the possibility of intentions about the actions *of others* - e.g. agent S1 intends agent S2 to *φ* - and the reducibility of these intentions to own-action intention (Campbell 2019, Ferrero 2013). I ignore this debate here.

fundamental units of theoretical thinking" (Castaneda 1971, p.465). Crucially, these units contain an agent and an action-type. Chisholm takes intention-contents to include a generic action-type and the state of affairs aimed at by the agent. This sort of content is captured by the formulation: S intends to bring it about [generic act-type] that *p* [pursued state of affairs] (Chisholm 1970). As far as I know, the only theory that puts neither action nor agent in the content of intentions is that of Ferrero, who takes intention-contents to be propositions representing the state of affairs that the agent pursues or 'aims at' - e.g. intending that *the temperature is above 20 degrees* (Ferrero 2013). This theory clashes with the surface grammar of intention-attributions, but it is the only theory that allows speakers to intend *that demonstrative d refers to object o in context c*, which is close enough to the content of (SI). One must then either endorse Ferrero's unorthodox theory of intention-content to uphold (SI), or abandon (SI) and put forward a content for semantic intentions that is consistent with mainstream theories of intention-content. The latter option seems more reasonable, but let me first argue that the former option might prove difficult to develop in conjunction with semantic intentionalism.

Ferrero treats intentions as aimings: one aims at a state of affairs. Aiming-attitudes 'move' the agent towards the pursued state of affairs: the agent adapts his behaviour so that the state of affairs obtains. On the face of it, this does not distinguish aimings from desires, but Ferrero thinks that these two attitudes can be distinguished through different conditions of success (Ferrero 2013, p.73). Now, if intentions about reference act as a determinant of the reference-state-of-affairs aimed at by the speaker, there is at least one thing the speaker can do in addition to uttering the relevant expression, and that is to form the right intention. But if this intention is the sole determinant of reference, the relation between intention and intended state of affairs is unlike that envisaged by Ferrero. The reference-state-of-affairs obtains just by uttering an expression with the intention that the reference-state-of-affairs obtains. There is no sense in which the intention 'moves' the agent towards her aim, since acting with this intention is enough to achieve the aim. Alternatively, if semantic intentions are not the sole determinant of reference, their 'moving' power seems to depend on whether additional determinants of reference can guide the speaker's behaviour - i.e. whether the speaker can do something to bring about these additional determinants. *Salience* seems to do well on this front for example: the speaker can do something to make an object salient - e.g. point in its direction - and thus aiming at *object o being salient* can move the speaker to act in ways that will not only achieve this intermediary aim, but also achieve the further aim that *d* refers to *o*. Other putative additional determinants of reference do not fare as well. King thinks that in addition to the speaker having an intention about *d* referring to o, it is necessary that an ideal audience with certain properties would take *d* to refer to *o* (King 2014). It is not clear that this modal state of affairs could guide the agent's behaviour in the way envisaged by Ferrero. This short discussion must remain inconclusive, but its point is simply that the only theory of intention-content consistent with (SI) is not a straightforward match for semantic intentionalism.

The more reasonable option for semantic intentionalists is to find another content for semantic intentions, one that is consistent with mainstream theories of intention-content. Some

alternative contents put forward in the literature do not fare better on this front. Gauker characterises the relevant intention as follows:

> (SI*) The speaker intends to utter demonstrative *d* in context *c* in such a way that *d* refers to *o* in *c* (adapted from Gauker 2018, p.118).[36]

This intention-characterisation is of the form 'S intends to φ in such a way that *p*', where φ is an act-type and *p* is the pursued state of affairs. This form does not correspond to any of the theories of intention-content I have reviewed. But it has the merit of matching some pre-theoretical attributions of intentions, e.g. 'The football player intends to kick the ball in such a way that it goes around the wall and lodges itself in the top-right corner of the net'.

An independent issue with (SI*) is the adverbial construction 'in such a way that…'. In what way does the speaker intend to utter demonstrative *d*? One might answer: in a way such that *d* refers to *o* in *c*. But there is something unsatisfying about this answer. Say I ask a football player how she plans to put the ball in the top-right corner of the net, and she answers that she plans to kick the ball in such a way that the ball reaches the top-right corner of the net. She has not really answered my question: what she intends to do in order to reach her goal remains unspecified. The same could be said of the speaker of (SI*): what she intends to do in order to make *d* refer to *o* remains unspecified.

Now, an agent cannot initiate her action on the basis of an intention to perform an unspecified act. Picture our football player planning her free-kick. If she merely intends to kick the ball *in such a way that it reaches the top-right corner of the net*, how can she initiate her run towards the ball? She must decide on a particular way to kick the ball. An intention to φ in such a way that *p* (that the pursued state of affairs obtains) may have its utility in action-planning, but further intentions must be formed in order for action to be initiated. In the case of reference-determining intentions, one would hope that a reference-determining intention is a downstream, action-initiating intention - an intention the speaker can act upon.

Having said this, an action-initiating intention downstream of (SI*) might not be suitable to determine reference either. Our football example can help to understand this. The football player might believe that brushing the bottom right side of the ball with the inside of her right foot will give the ball the right trajectory to reach the top-right corner of the net. And so she might form the intention to kick the ball in this particular way. If she does kick the ball in this way and the ball does reach the top-right corner of the net, this success will be due to the particular way in which she kicked the ball. Analogously, if the speaker intends to utter *d* in way *w* so that *d* refers to *o* in *c*, it seems that it is the way *w* of uttering *d* that allows *d* to refer to *o* in *c*, rather than the intention to utter *d* in way *w*. A committed semantic intentionalist could always reply that the way *w* to utter *d* so that *d* refers to *o* in *c* is to utter *d* with the right

[36] Gauker actually writes "in such a way, in such circumstances, that…". Gauker takes *utterances of expressions,* rather than *expression-types in context,* to refer to objects. 'In such circumstances' seems to be an oblique appeal to the context of utterance, which can be removed if an *expression-types-in-context* view of the bearers of reference is assumed.

semantic intention. But then the semantic intention would be reflexive: it would be an intention to utter *d with this very intention* so that *d* refers to *o* in *c*. I will not settle here whether reflexive intentions should be countenanced, but suffice to say that reflexivity is bound to be a controversial feature.[37] In summary, the 'in such a way' qualification in (SI*) raises several worries in addition to (SI*) not fitting extant theories of intention-content.

Another author who speaks of 'ways of using' expressions when discussing semantic intentionalism is Viebahn (Viebahn 2020). Viebahn thinks it plausible that speakers have intentions about the 'semantic value' (i.e. reference, or content more generally) of the expressions they utter, although these intentions are not best conceived as intentions *that something be the semantic value of an expression*. For Viebahn, uttering an expression in a certain way is uttering it *with certain properties it possesses*. Reference-determining semantic intentions could then take the following form:

$(SI^{**})_1$ S intends to utter demonstrative *d* with reference *o* in *c*.

If one prefers a propositional view of intention-content, the intention can be formulated as:

$(SI^{**})_2$ S intends that S utters demonstrative *d* with reference *o* in *c*.

Viebahn's idea yields contents for intentions about reference that are consistent with either the propositional view of intention-content or the act-type view. In conclusion, although some proponents of intentions about reference have been careless with the content of these intentions, a *prima facie* acceptable content can be found in the literature.

2. Do speakers have intentions about reference?

Intentions about semantic reference are posited by intentionalists who do not fancy communicative intentions as determinants of semantic reference. But are there independent reasons to think that speakers have intentions about the reference of the expressions they utter? Bach thinks not, given that speakers already have the communicative intentions constitutive of speaker reference:

> Defenders of semantic intentionalism about demonstratives simply assume that we have intentions that determine semantic values (references) of uses of demonstratives in contexts. To me it is anything but obvious that we do. I don't know about you, but when I use a demonstrative, the only intention I have that is directly associated with the demonstrative is to refer to something. (Bach 2017, p.72)

[37] Some authors think that *communicative* intentions are reflexive (Bach & Harnish 1979). This feature troubles other authors (Recanati 1986, Sperber & Wilson 1995).

Of course, if one is a sceptic about communicative intentions, the dialectic looks different. But here, following Bach, I will assume that speakers do have the communicative intentions constitutive of speaker reference. Bach then challenges semantic intentionalists to provide independent reasons to posit intentions about the reference of expressions on top of the communicative intentions constitutive of speaker reference.

a. Viebahn's case from encoding intentions

Are there independent reasons to think that speakers have the following sort of intention?

$(SI^{**})_1$ S intends to utter demonstrative *d* with its reference *o* in *c*.[38]

Viebahn takes up Bach's challenge in (Viebahn 2020). He accepts the existence of communicative intentions, and yet tries to show that speakers also have intentions about the reference of the expressions they utter. Viebahn posits several varieties of intentions about reference, but here I will focus on what he calls 'encoding' intentions.[39] I take encoding intentions to constitute a subcategory of intentions of the form 'S intends to φ by ψ-*ing*', where φ is the content of a communicative intention and ψ is a way of φ-ing. I take the liberty to call intentions of this form *implementation intentions*. Implementation intentions are intentions to implement a communicative intention in a certain way. An encoding intention is then a species of implementing intention. The encoding intention that is crucial for Viebahn's purposes is the following:

> The speaker intends to communicate that *p* by uttering a sentence S with content *p* in context of utterance *c*.

The terminology 'encoding' is transparent: the speaker intends to communicate a proposition by using a linguistic vehicle that encodes this proposition. However, it seems that Viebahn uses 'encoding intention' to designate not the complex intention above, but rather its

[38] Here is a naive justification for the postulation of $(SI^{**})_1$. Given an intention to utter *e* and a belief that expression *e* refers to *o* in *c,* the speaker must also have an intention to utter *e* with its reference *o* in *c*. Unfortunately, this justification is not sound if it is put forward as an instance of a general closure principle about intentions. I may intend to cut onion *o* and believe/know that *o* has a weird shape, yet I do not additionally intend *to cut o with its weird shape*. A version of the underlying closure principle restricted to linguistic expressions and their properties fails too. It is not true that: whenever I intentionally utter expression *e* and know that *e* has property Q, I further intend to utter *e* with Q. Not all that one knows or believes about the expressions one utters figures in one's intentions in uttering the expression.

[39] Viebahn also posits 'equivalence' and 'deferential' intentions, which cover cases in which speakers intend to use expressions in the same way that other expressions are used, or in the same way as other speakers, respectively (Viebahn 2020). As explained by Viebahn, the words of these speakers intuitively express a content, and thus semantic intentionalists must attribute appropriate semantic intentions to them too. 'Initial' uses of demonstratives are the ones that dominate the metasemantics literature, and equivalence or deferential intentions would not help with these uses. If Viebahn is right that speakers have equivalence and deferential intentions (I will not investigate this here), the negative conclusion of this chapter should be restricted to the semantic intentions needed by semantic intentionalists to account for initial uses of demonstratives.

component [The speaker intends to utter S with its content *p* in *c*] (Viebahn 2020, p.98). At the risk of sounding fussy, I prefer to reserve 'encoding intention' for the complex intention. Only with the complex intention do we get the idea of an intention to convey a message by using a vehicle that encodes the message. On its own, the component intention *to utter S with content p in c* is not in any sense an intention to encode a message in a vehicle. One could intend to communicate that *q* by uttering S with content *p* in *c*, and thus intend t*o utter S with content p in c*, yet there would be no intention to encode one's intended message in linguistic vehicle S.

Viebahn takes the component intention *to utter sentence S with content p in c* to determine the content of S in *c*. But he also takes this component intention to determine the content of the *sub-sentential expressions* contained in sentence S. For example, the intention *to utter sentence S with content p in c* may determine the reference *o* of a sub-sentential expression contained in S (Viebahn 2020, p.106). The alternative here is to allow speakers to have further intentions about the content of sub-sentential expressions, and to let these further intentions determine the content of sub-sentential expressions. Viebahn has his reasons for preferring the first option, reasons which I cannot review here. In what follows I freely move between intentions about the content of sentences and intentions about the content of sub-sentential expressions, assuming that they could equally play the reference-determining role.[40]

Viebahn's view is then that intentions about reference can be derived from encoding intentions. My first problem with this view is the following: if there are implementing intentions of the encoding sort, then they are not quite the intentions postulated by Viebahn. Say that a speaker intends to refer to object *o*, intends to do so linguistically, and intends to do so linguistically in the most straightforward way possible: by uttering a linguistic expression that refers to *o* in the context of utterance. The speaker's overall intention can be formulated as:

> The speaker intends to refer to *o* by uttering *some expression (or other)* that refers to *o* in *c*.[41]

---

[40] So technically Viebahn does not posit intentions about the reference of demonstratives. This is awkward, but Viebahn's general policy regarding intentions about the content of sub-sentential expressions has nothing to do with intentions about the reference of demonstratives. He rather follows Glanzberg in being wary of intentions about arcane content-entities such as functions: it seems implausible that speakers have intentions about functions (Glanzberg 2007; Viebahn 2020, pp.103-104).

[41] I do not know how to conceive of the logical form of this intention-attribution. It seems to require quantification over expressions. But a quantifier with scope over 'S intends' gives the wrong result (*de re* intentions about expressions), while quantification under the scope of 'S intends' gives awkward results given mainstream theories of intention-content. For instance with existential quantification: S intends that [there is some expression *x* that refers to *o* and S refers to *o* by uttering *x]*. This problem seems to be an instance of the more general problem of the interaction of quantification with non-doxastic attitude verbs (Schoubye 2011). In addition to this difficulty, the 'by' construction that comes with complex intentional action raises another difficulty: what is the logical form of 'S *φ's* by *ψ*-ing'?

We can call this intention a *generic* encoding intention, as opposed to the *singular* encoding intention about a particular expression postulated by Viebahn, to wit:

> The speaker intends to refer to *o* by uttering expression *e* with reference *o* in *c*.

In a generic encoding intention we find a component intention *to utter some expression (or other) that refers to o in c*. We can call this latter intention a *generic semantic intention*. This generic semantic intention is not the intention that is supposed to determine reference according to Viebahn. *That* intention is a singular intention about the particular expression that the speaker intends to utter.

The question we face is then the following: does having a generic encoding intention entail having a singular encoding intention? It seems to me that Viebahn simply assumes a positive answer:

> … Anne forms the communicative intention to tell Bert that he can have the book that is lying on the table in front of them. [...] One way in which Anne might intend to achieve her communicative goal is by uttering expressions that encode or express (or in more technical terms: have as their semantic value) what she intends to communicate. So we might say that Anne intends to communicate that Bert can have the book that is lying on the table in front of them by encoding her message in the string of expressions 'You can have that'. (Viebahn 2020, p.98)

The intention in the second sentence of the quoted passage is not about a specific string of expressions: Anne intends to inform Bert that *p* by uttering some string of expressions (or other) that encodes *p*. This is a generic encoding intention. Because the sentence uttered by Anne - 'You can have that' - has already been specified prior to the quoted passage,Viebahn seamlessly attributes in the third sentence of the passage a further intention to Anne: the intention to inform Bert that *p* by encoding *p* in 'You can have that'. Hence, Viebahn moves without warning from a generic encoding intention to a singular encoding intention.

This transition must be justified. To assess whether generic encoding intentions necessarily come with singular encoding intentions, it might help to provide a rationalising explanation of a speaker's action, assuming that this speaker has a generic encoding intention. Does a rationalising explanation require the further attribution of a singular encoding intention? Take Anne's case. Anne intends to inform Bert that *p*. She wants to do this linguistically, in the most straightforward way possible. She thus intends to utter some string of expressions that has *p* as its content in the context of utterance (generic encoding intention). She believes that 'You can have that' can have *p* as its content in the context, given the metasemantics of 'That'. We can here assume that the metasemantics of 'That' is encoded in its linguistic meaning, and thus that Anne's knowledge of the metasemantics of 'That' is part of her linguistic knowledge.

At this dialectical juncture, one cannot assume that semantic intentionalism is the true metasemantics of 'That'. Indeed, Viebahn wants to provide an *independent reason* to postulate intentions about reference; that is, a reason independent from wanting non-communicative intentions to determine reference. One must then establish that these intentions about reference exist whether or not they determine reference. Back to our rationalising explanation, suppose for the sake of illustration that the right metasemantics for demonstratives is a *demonstration* metasemantics. The referent of 'That' in a context of utterance is the object determined by the ostensive gesture performed by the speaker. Anne knows this, and so she forms the intention to utter 'You can have that' while pointing at the book on the table. Acting in this way will be a way to implement her generic encoding intention. Overall, Anne intends to inform Bert that *p* by uttering 'You can have that' while pointing at *o* - the book on the table. This rationalising explanation does not include an intention *to inform Bert that p by uttering 'You can have that' with content p in c*. If this putative intention is an idle wheel in a rationalising explanation of Anne's action, why should one posit it?

More anecdotally, even if one supposes that semantic intentionalism is the true metasemantics for 'That', and thus *presupposes* that singular semantic intentions exist, one does not quite get singular encoding intentions out of generic encoding intentions. If semantic intentionalism is the true metasemantics for 'That', then Anne knows that: 'You can have that' will have *p* as its content in the context of utterance *provided she forms the right singular semantic intention*. Anne thus forms the intention to utter 'You can have that' with the right singular semantic intention. Overall, Anne forms the intention to tell Bert that *p* by uttering 'You can have that' with the intention to utter 'You can have that' with content *p* in *c*. Let me put next to each other the intention postulated by Viebahn and the intention reached in this paragraph:

> Anne intends to inform Bert that *p* by uttering 'You can have that' with content *p* in *c*.
>
> Anne intends to inform Bert that *p* by uttering 'You can have that' with the intention to utter 'You can have that' with content *p* in *c*.

In summary, Viebahn's case for singular semantic intentions is not compelling. It is plausible that speakers have *generic* encoding intentions, but speakers must further have *singular* encoding intentions in order to have singular semantic intentions. And I see no reason to think that speakers with generic encoding intentions also have singular encoding intentions. On the contrary, a rationalising explanation of the action of a speaker who has a generic encoding intention seems to leave no room for a singular encoding intention.

b. Viebahn's case from polysemy

When a speaker utters 'I drank the whole bottle', it is clear that the operative meaning of 'bottle' is [bottle-content] rather than [liquids-container with tapered neck]. Viebahn claims,

following Bach, that the operative meaning of 'bottle' in context is determined by the speaker's intention to use 'bottle' with the meaning [bottle-content], and thus claims that we have here an uncontroversial case of a semantic intention determining a semantic property. If speakers have intentions about the meaning of polysemous expressions, Viebahn continues, there is no good reason to deny that they have intentions about the reference of demonstratives. Hence, Bach should not resist intentions about the reference of demonstratives.

I do not hold a strong view on the conditional put forward by Viebahn - i.e. that if speakers have intentions about the meaning of polysemous words, then they also have intentions about the reference of demonstratives. Let me just draw a contrast between these intentions that might give one pause. Intentions about the meaning of an expression - e.g. an intention to use expression *e* with its meaning *m* - are intentions about a property that the expression possesses independently of the context of utterance. An expression *e* has meaning *m* in language L whatever the context of its utterance. A fortiori, the fact that *e* has meaning *m* is independent from the speaker's intention to use *e* with meaning *m*. By contrast, an intention about the reference of a demonstrative in context is an intention about a context-dependent property. If we further assume that an intention about (context-independent) meaning determines (context-dependent) operative meaning on the one hand, and that an intention about (context-dependent) reference determines (context-dependent) reference on the other hand, we get quite different determination-structures. In the case of polysemous expressions, an intention about property P_1 determines a *distinct* property P_2. In the case of the reference of demonstratives, an intention about property P determines *this very property P*. The polysemy case and the reference case are dissimilar in this respect, although I admit I cannot see through the consequences of this difference.

My main concern is not with the conditional relating meaning-intentions and reference-intentions, but rather with the claims about polysemy that Viebahn endorses. First of all, it is not entirely uncontroversial that polysemy is a semantic phenomenon. As explained in (Carston 2021), one view about polysemy is that the lexically encoded meaning of polysemous expressions is a very thin schema, while the operative meaning - the sense that the expression gets when uttered by a speaker - is always a pragmatic property of speakers. This view is not the most popular one, but it is a first hurdle for the line of thought pushed by Viebahn.

Secondly, even if we grant that operative meaning in context is a semantic property, an intentionalist metasemantics for this property is not obvious. Although what the speaker means with 'bottle' is clearly a matter of her intentions, it is not obvious that the operative meaning of 'bottle' in a context is a matter of her intentions. Viebahn acknowledges this, but he makes it sound as if intentionalism about operative meaning in context is the overwhelmingly plausible view to take (Viebahn 2020, pp.100-101). It seems to me that other views are just as plausible. The operative meaning of a polysemous word in context could be the most salient one in the context of utterance, rather than the one intended by the speaker. This 'metasemantics' of operative meaning would give the right result in the present case:

[bottle-content] is the most salient meaning because ‘bottle’ occurs as the direct object ‘drink’ in the sentence uttered by the speaker. The operative meaning of a polysemous expression in context could alternatively be the one that a competent, attentive, and reasonable interpreter would retrieve. Again, this would give the right result in the present case. In other words, there are plausible non-intentionalist metasemantics for operative meaning in context.

Now, the truth of intentionalism about operative meaning is not crucial to Viebahn’s argument from polysemy. Viebahn wants to argue that the existence of intentions about the meaning of polysemous expressions makes the existence of intentions about content plausible. If there are independent reasons to think that speakers have intentions about the meaning of polysemous words (independent from the issue of the determination of operative meaning), then the core of Viebahn’s argument is safe. As it turns out, Viebahn does attempt to provide an independent reason to posit intentions about the meaning of polysemous words. He claims that these intentions are embedded in certain implementation intentions (Viebahn 2020, p.102). In our ‘bottle’ case, the speaker intends to communicate a proposition about a bottle-content by using ‘bottle’ with its [bottle-content] meaning.

Again, I would reply that speakers do not have singular implementation intentions of this sort, for the same reason that they do not have singular encoding intentions. A rationalising explanation of the speaker’s utterance of ‘I drank the whole bottle’ would not include an intention *to communicate something about a bottle-content by using ‘bottle’ with its [bottle-content] meaning*. The speaker intends to communicate that she drank the whole content of the bottle of kombucha that was in her fridge - that *p*. She wants to communicate this linguistically, and as straightforwardly as possible. So she forms the intention to communicate that *p* by uttering a sentence that has *p* as its content (generic encoding intention). She knows that ‘bottle’ has [bottle-content] as one of its meanings, and thus that uttering ‘bottle’ allows her to implement her generic encoding intention. She further knows the determinants of operative meaning, and plans her utterance accordingly. As in the previous subsection, given a non-intentionalist metasemantics for operative meaning, we do not get the singular implementing intention postulated by Viebahn.

Here is a final consideration against the idea that speakers have intentions *to utter polysemous expression e with its meaning* m_1. These intentions make it look as if the speaker *selects* one meaning among the several meanings of the polysemous expression she utters. But intuitively, that’s just not how things go. The speaker does not choose a polysemous expression and then picks one of its meanings. She rather intends to communicate a content, and then picks an expression that has a meaning which allows the expression of her intended content. The speaker wants to communicate a proposition about the content of a bottle, and believes (correctly) that ‘bottle’ can express this content in the context of her utterance, so she decides to utter ‘bottle’. She does not select one meaning among the several meanings of ‘bottle’. Meaning-selection is a task for *the interpreter* facing the utterance of a polysemous word, not for the speaker who uses it.

In this subsection I have argued as if Viebahn himself holds all these views about polysemy which I do not find compelling. This is somewhat misleading. When defending intentions about reference, Viebahn's interlocutor is Bach. Bach accepts that the operative meaning of polysemous words in context is fixed by intentions of the speaker about the meaning of these words (Viebahn 2020, p.100). So Viebahn's argument from polysemy is that Bach should welcome intentions about reference given his appreciation of intentions about the meaning of polysemous words. Be that as it may, if the case for speaker intentions about the meaning of polysemous words is not sound, a case for intentions about reference that rests on a case for intentions about the meaning of polysemous words is not sound either. Persuading Bach to endorse intentions about reference is one thing; persuading everyone else is another.

c. Communicative intentions in disguise?

In this subsection I assume that intentions about reference are best formulated as intentions *to utter an expression e with reference o in context c*, along the lines recommended by Viebahn. However, it must be noted that any intention-content fit to determine the reference of *e* would do. To be fit to determine reference, the intention-content must mention the relation between expression *e*, object *o*, and context *c* (or between an utterance of *e* and object *o*). In other words, the intention-content must be fully singular. An intention *to utter some expression that refers to o in c* would not do for example, because it could not determine the reference of a particular expression *e*. An intention *to utter e with its reference* (whatever that is) *in c* would not do either, because it would not target any particular object. By contrast, an intention *to utter e so that e refers to o in c* would do just as well as an intention *to utter e with reference o in c*.

I want to argue that in the absence of a distinct mechanism for the recognition of singular intentions about reference, there is no reason to posit such intentions in addition to communicative intentions. Let me first explain what I take to be the task of an interpreter faced with the utterance of a demonstrative under semantic intentionalism. I assume that the metasemantics of demonstratives is linguistically encoded: the meaning of demonstratives tells us in what conditions demonstratives refer to a certain object in a context of utterance. If semantic intentionalism is the true metasemantics for demonstratives, then the meaning of demonstratives tells us that demonstrative *d* refers to *o* in *c* only if the speaker intends to utter *d* with reference *o* in *c*. The interpreter automatically applies this linguistic rule whenever faced with an utterance of *d*. The interpreter's task is then to recognise the object targeted by the speaker's semantic intention. This task may be seen as as that of filling the blank in the following intention:

> The speaker intends to use *d* with reference _ in *c*.

Now, how does the interpreter fulfil this task? Unless a special intention-recognition mechanism is postulated that is not involved in the recognition of communicative intentions, our default-view should be that mechanisms involved in the recognition of communicative

intentions are recruited for the task at hand. But if mechanisms for the recognition of *communicative* intentions are recruited to recognise the object targeted by the speaker's *semantic* intention, then it seems that recognising the object targeted by the speaker's semantic intention is just recognising the speaker's referent. If this is the case, it is unclear why one would maintain that the intention to be recognised is *distinct from the speaker's communicative intention*. The speaker expects the interpreter to recognise her allegedly semantic intention by employing mechanisms dedicated to the recognition of communicative intentions. The interpreter attempts to recognise the allegedly semantic intention by employing these mechanisms. Why then would one contend that the allegedly semantic intention is anything other than a communicative intention?

I expect several replies to this line of thought. The first reply is that I have ignored the possibility that there are non-intentional determinants of reference in addition to semantic intentions. If there are such additional determinants, then ascertaining semantic reference cannot just be a matter of ascertaining speaker reference. I see nothing wrong with this reply, but it is not an objection to my main claim. My main claim is that recognising the object targeted by a speaker's singular semantic intention involves the same mechanisms as recognising the speaker's referent, barring reasons to think that a special, distinctive mechanism is involved. This leaves open the possibility that ascertaining semantic reference demands more than ascertaining speaker reference, if there are additional non-intentional determinants of reference.

The second reply I expect is that I have ignored the possibility that the metasemantics of demonstratives are *not* linguistically encoded, which would free the interpreter from having to recognise the speaker's singular semantic intention. On this view, reference is determined by the speaker's singular semantic intention, but this determination is not linguistically encoded, and thus not known by competent speakers as a matter of linguistic knowledge. The question raised by this view is whether competent speakers know in another way what determines the reference of demonstratives. If they do, then they are still saddled with the task of recognising the speaker's semantic intention, and my main claim stands. Now, what if competent speakers do not know what determines reference? Åkerman distinguishes between metaphysical and epistemic determinants of reference (Åkerman 2009). According to Åkerman, the contextual facts that determine reference - in the metaphysical sense of 'determine' - need not be facts that interpreters have access to in determining reference - in the epistemic sense of 'determine'. Interpreters might ascertain the reference of an expression with no access to the contextual facts that metaphysically determine it. To be clear, Åkerman still assumes that interpreters have access to *reference facts* themselves. The question is only which facts give interpreters access to reference facts. Of course, interpreters has access to various contextual facts: what objects were referred to in previous discourse, what objects are perceptually salient, what objects in the environment satisfy the predicate of the sentence uttered by the speaker, what objects are located in the direction of the speaker's ostensive gesture, etc… The facts through which interpreters ascertain reference facts could be one of these, or any combination of these. But notice that the contextual facts mentioned here are *broad* contextual facts, i.e. facts that help the interpreter ascertain speaker reference (see section 2 of the

previous chapter). In addition, interpreters faced with the utterance of a demonstrative still need to ascertain speaker reference, whether or not the determinants of semantic reference are linguistically encoded. Therefore, the distinction made by Åkerman does not help avoid the conclusion that ascertaining speaker reference, either directly or by proxy through wide contextual facts, is involved in ascertaining semantic reference. Why then would one maintain that semantic reference is metaphysically determined by epistemically inaccessible semantic intentions, when one could just say that it is metaphysically determined by epistemically accessible communicative intentions?

The third reply I expect is that there is an independent reason to resist the idea that communicative intentions determine semantic reference. This reason is the necessity of preserving the distinction between speaker reference and semantic reference (Kripke 1977). Granting here the necessity of preserving this distinction, I would simply answer that taking communicative intentions to be determinants of semantic reference does not preclude a distinction between semantic reference and speaker reference. I made this clear in the previous chapter with the help of Neale-Schiffer-style metasemantics (Neale-Schiffer 2020):

> *x* is the referent of expression *e* on an occasion of use *o* if and only if:
> (1) C*x* (C is a constraint set by the meaning of expression *e*)
> (2) the speaker refers to *x* with *e* on *o*

On this view, semantic reference is only partly determined by the communicative intentions that constitute speaker reference, and thus semantic reference is distinct from speaker reference. Of course, if one expects the speaker reference / semantic reference distinction to make it possible that the speaker referent is o_1 while the semantic referent is distinct object o_2 - that speaker reference and semantic reference diverge - having communicative intentions as determinants of semantic reference is not an option. And there might be good reasons to expect this: the intuitive possibility of a speaker saying something about an object and meaning something about a *different* object, combined with the general assumption that say-judgements track semantic reference, would justify resisting reference-determining communicative intentions. In any case, the desideratum of possible divergence between semantic reference and speaker reference goes beyond the desideratum to preserve a distinction between semantic reference and speaker reference.[42] Additionally, satisfying the distinction desideratum *and* the divergence desideratum while holding that speaker reference and semantic reference are both determined by communicative intentions is possible. If one takes communicative intentions to be structured - i.e. intentions to φ by ψ-ing - one can contend that one component of communicative intentions determines speaker reference (the intention to φ), while another component determines semantic reference (the intention to ψ) (Bach 1992).

[42] Viebahn seems to want the stronger desideratum to constrain a metasemantics for demonstratives, and seemingly for the reasons given in the main text, but he does not make explicit all the assumptions that come with this desideratum (Viebahn 2020, pp.96-97).

Let me return to my initial argument. It may be taken as a challenge for semantic intentionalism. If the recognition of singular semantic intentions involves the mechanisms also involved in the recognition of communicative intentions, there is no reason to postulate the former intentions in addition to the latter. So semantic intentionalists should put forward a distinct mechanism for the recognition of their favoured intentions.

Now, a semantic intentionalist who also happens to be a sceptic about communicative intentions might reply that she is not concerned with this challenge. Indeed, it seems that if one does not believe in communicative intentions in the first place, one is not concerned with distinguishing semantic intentions from communicative intentions. However, the challenge I present does not ultimately depend on the existence of communicative intentions. It just depends on the existence of speaker *reference*. My challenge could be reformulated as follows: barring reasons to think that a distinctive mechanism is at play in filling the blank in [The speaker intends to use *d* with reference _ in *c*], the default view should be that the mechanism involved in ascertaining *speaker reference* is employed.

Conclusion

Semantic intentionalism faces two conceptual problems. First, most of its proponents saddle intentions about reference with implausible contents. Fortunately, a more plausible content can be found in the literature (section 1). Secondly, semantic intentionalists must provide independent reasons for the postulation of intentions about reference. Viebahn's cases from encoding intentions and polysemy are unpersuasive (sections 2.a. and 2.b.). In addition, without the postulation of a recognition-mechanism distinct from that involved in the recognition of communicative intentions, the postulation of intentions distinct from communicative intentions is unwarranted (section 2.c.).

Chapter 4: Classical and non-classical metasemantics

So far I have argued against a prominent family of metasemantics for demonstratives called *intentionalism*. In this fourth chapter I want to reflect on metasemantics for demonstratives more generally. Here are three possible desiderata for a metasemantic for demonstratives:

(1) Provide individually necessary and jointly sufficient conditions for an object to be the referent of a demonstrative in context.
(2) Double as the linguistic meaning of demonstratives: the determinants of reference are encoded in a linguistic rule mastered by speakers of the relevant language.
(3) Make predictions about reference that match pre-theoretical say-judgements on cases (e.g. the judgement that the speaker said something about object *o*).

This package of theoretical desiderata defines what I would like to call *classical metasemantics* for demonstratives. The package is implicitly or explicitly endorsed in much of the work following Kaplan's pioneering contribution (Kaplan 1989a). Here is a non-exhaustive list of authors who endorse the three above desiderata, and thus qualify as classical metasemanticists: King (2013; 2014), McGinn (1981), Speaks (2017), Stokke (2010), and Wettstein (1984).

As we will shortly see, many theories fail *qua* classical metasemantics. We already know that various forms of intentionalism fail to satisfy one of desiderata (1)-(3). In Chapter 1 we saw that both a naive form of intentionalism and a more sophisticated form of intentionalism called *proximal intentions intentionalism* fall foul of desideratum (3). In Chapter 2 we saw that communicative intentionalism is not plausible if desideratum (2) is taken on board. So why not go for a *non-classical metasemantics*? That is, why not reject some of the three desiderata? The hope is that going non-classical will rejuvenate old metasemantics, and perhaps make room for novel metasemantics.

In section 1 I present some aspects of Kaplan's views on demonstratives (Kaplan 1989a, 1989b). This partial exposition serves as a genealogy of the three desiderata that define classical metasemantics. In section 2 I lay out the case for going non-classical: many classical metasemantics fail; and there are reasons to think that desiderata (2) and (3) cannot be jointly satisfied. In the remainder of the chapter, I explore the consequences of rejecting each desideratum of classical metasemantics in turn: desideratum (1) in the third section; desideratum (2) in the fourth section; and desideratum (3) in the fifth and final section.

1. From Kaplan to classical metasemantics for demonstratives

One aim pursued by Kaplan in *Demonstratives* (1989a) is to capture pre-theoretical platitudes about expressions such as 'I' and 'This' in a compositional, truth-conditional theory of meaning - a CT theory for short. Here is a first platitude. With expressions such as 'I' and 'This', different objects are spoken of in different contexts. When you utter 'I am tired', you say something about yourself. When I utter 'I am tired', I say something about myself. When I utter 'This is pretty' pointing at my pink mug, I say something about my pink mug. When I utter 'This is pretty' pointing at my blue mug, I say something about my blue mug.

As for the second platitude, here is something English speakers know about 'I': whenever a speaker utters 'I', that speaker says something about herself.[43] In some intuitive sense, this general principle gives *the meaning* of 'I'. Hence Kaplan aims to capture in a CT theory the two obvious facts that: (a) the object about which something is said with 'I' is different in different contexts, although (b) the meaning of the word 'I' remains the same across contexts. Here is a final fact about 'I', though I am not sure it qualifies as a pre-theoretical platitude. What is said with a sentence containing 'I' is different in different contexts *because of* the meaning of 'I'. Facts (a) and (b) are not independent. The context-invariant meaning *determines* the context-variant object about which something is said. Kaplan aims to capture this third fact.

Let me move on to *how* Kaplan captures these facts in a CT theory. Here is a crucial thesis in Kaplan's theory. What a speaker says with e.g. 'I am tired' or 'This is beautiful' is about a particular object in the following sense: the contribution of 'I' and 'This' to what the speaker says *is* this object. Kaplan uses the term 'content' for both what is said with a sentence and for the contribution to what is said of a sub-sentential expression. We might then call Kaplan's thesis *singular content*. Kaplan himself calls it *direct reference*, but I will explain in a second why I prefer to reserve this terminology for another thesis.[44]

Kaplan assumes that what is said - content - can be captured by a CT theory. If 'reference' and its cognates designate a relation between linguistic expression and object that obtains in a CT theory, one aspect of Kaplan's project is the following identification: the content of a linguistic expression = its *referent*. Desideratum (3) of classical metasemantics follows trivially. A metasemantics should make predictions about reference that match pre-theoretical

---

[43] Kaplan does not think that the bearers of reference are *utterances of expressions*. He rather takes *expression-types in context* to be the bearers of reference. For coherence, the meaning of 'I' should not mention *utterances* or *utterers* given an expression-type-in-context view of the bearer of reference. Kaplan allows himself to slip into 'utterance' talk in the less technical parts of his exposition (Kaplan 1989a, p.520).

[44] Kaplan calls *singular content* ('direct reference' in his own words) an "obvious principle" (1989a, p.492). This is slightly confusing, as Kaplan spends some time arguing for *singular content*. As I see it, *singular content* is obvious in the sense that it can be motivated with pre-theoretical judgements on sentences only (Kaplan 1989a, section V and X). After Kaplan's *Demonstratives*, several theorists have presented sentences for which our pre-theoretical judgements tell against *singular content* (Elbourne 2008, King 2001, Roberts 2002).

say-judgements for the simple reason that what the speaker says (content) is explained in terms of reference.

Kaplan thus wants a CT theory to capture the fact that the content of 'I' is the object about which something is said. The most straightforward way to do so is to identify the compositional value of 'I' - i.e. the entity associated with the expression and entering composition in a CT theory - with this object. This straightforward way is Kaplan's preferred way. In general, Kaplan wishes to identify the compositional values of expressions in a CT theory with their content. In particular, the object about which something is said with 'I' is identified with the compositional value of 'I'. The CT theory thus captures the relation between linguistic expression and object *directly*, by making the object the compositional value of the expression. It is for this aspect of Kaplan's view that I reserve the phrase 'direct reference'.[45]

Recently, some authors have insisted that entering the metasemantics for demonstratives debate does not require endorsing *direct reference* (King 2014, Speaks 2016, Viebahn 2020). One can hold that demonstratives refer without holding that their compositional value is the object about which something is said. The consensus seems to be that one can speak of a relation of reference (and of its determinants) as long as the compositional value posited for demonstratives *determines* the object about which something is said.[46]

Let me now turn to how Kaplan handles the meaning of e.g. 'I' and 'This'. In general, Kaplan thinks of the context-invariant meaning of expressions as *"linguistic rules known, explicitly or implicitly, by all competent users of the language"* (Kaplan 1989a, fn 13). In the case of words like 'I', the meaning is a rule stating what aspects of the context of utterance determine the reference of the word in the context. Such a rule is called by Kaplan a *character*. In set-theoretical terms, the character of 'I' is a function whose domain is the set of contexts of utterance and whose co-domain is the set of objects. The character of 'I' is {<x,y> : y is the speaker of x} where x is a context of utterance and y is an object. With Kaplan's characters, we get the idea of a linguistic rule that encodes the contextual determinants of reference, as per desideratum (2) of classical metasemantics.

I have traced back the origin of desiderata (2) in Kaplan's work. There is however a complication regarding the genealogy of (2). Kaplan draws a contrast between *pure*

[45] Notoriously, Kaplan's own logic for demonstratives violates *direct reference*. In his logic, compositional values are called 'contents', and yet for expressions such as 'I' and 'This' they are not the objects about which something is said. They are rather (constant) functions from worlds and times to objects.

[46] 'Determine the object about which something is said' can mean different things here. For King (2014), who has in mind a quantificational analysis of demonstratives, and thus gives a higher-order property as the compositional value of demonstratives, it means that the compositional value is satisfied by the object about which something is said. King and other metasemanticists use 'semantic value' instead of 'reference' to make it clear that they are not presupposing direct reference. See Rabern (2012) for another sense in which compositional value can determine content without being identical with it.

indexicals such as 'I' and *true demonstratives* such as 'This'. Pure indexicals have a character (1989a, p.491). They correspond to my *indexicals* of Chapter 2, expressions for which some contextual fact plays a *narrow* role. By contrast, true demonstratives have no character, despite having a context-variant reference. In Kaplan's own words:

> The linguistic rules which govern the use of the true demonstratives 'that', 'he', etc., are not sufficient to determine their referent in all contexts of use. Something else—an associated demonstration—must be provided. **The linguistic rules assume that such a demonstration accompanies each (demonstrative) use of a demonstrative** (1989a, p.490, emphasis mine).

This passage needs to be interpreted. First of all, for Kaplan a *demonstration* is not an ostensive gesture, although an ostensive gesture may determine a demonstration. Kaplan seems to think of demonstrations as non-linguistic equivalents of Fregean definite descriptions, since they are supposed to pick out objects *via* a mode of presentation (1989a, section IX). More accurately, demonstrations are the non-linguistic equivalents of a special sort of Fregean definite description. For Kaplan, the *same* demonstration can occur at different times and locations, and thus a demonstration is the non-linguistic equivalent of a Fregean description of the form 'The F here now' - e,g, *'The brightest heavenly body now visible from here'* (Kaplan 1989a, p.526). This individuation of demonstrations entails that a demonstration determines an object *in a context*.

Now turning to the sentence in bold in the quoted passage, if we take 'assume' at face value, the idea is that the meaning of 'That' or 'He' does not *state* that the referent is the object determined by a demonstration (in a context). Unfortunately, Kaplan does not offer a positive view of the meaning of true demonstratives. He only adds that true demonstratives, unlike pure indexicals, are *"incomplete expressions"*: they must be combined with a demonstration in order to obtain a character (1989a, p.527). Demonstrations may be regarded as the bearers of this character, since they determine an object in a context. In the logical form of sentences, demonstratives are augmented with a definite description representing the associated demonstration - e.g. *This [The F here now]*. And the compositional value of *This [The F here now]* is just the F here now, if there is such an object. The compositional value of a demonstrative is thus the object determined by its associated demonstration in context.

Kaplan calls this package of views about demonstratives the 'Corrected Fregean Theory'. As I understand this theory, demonstrations are not explicitly mentioned by the meaning of demonstratives. And yet demonstratives appear augmented with a demonstration in the logical form of sentences in which they occur. If this conjunction feels too uncomfortable, an easy remedy is to abandon the interpretation of 'assume' according to which linguistic rules are silent about demonstrations, while upholding the central claim that the meaning of demonstratives is not a character. Here is the resulting view. The meaning of a demonstrative *states* that the demonstrative refers to the object determined by the demonstration in context. In other words, the meaning of demonstratives is a function from demonstrations to

characters.[47] In this corrected Corrected Fregean Theory, the determinants of the reference of demonstratives - demonstration plus context of utterance - are encoded in the meaning of demonstratives, as per desideratum (2) of classical metasemantics.

Kaplan further presents an alternative to the Corrected Fregean Theory, which he calls the 'Indexical Theory'. In Kaplan's formal framework for both the Corrected Fregean Theory and the Indexical Theory, context is represented by a set of parameters including at least a time, a location, an agent, a world. The basic idea of the Indexical Theory is to move demonstrations from the logical form of sentences to the set of parameters that formally represents context.[48] A sequence of demonstrations can be introduced as an element of the context. Philosophically, if demonstrations are themselves features of the context of utterance, they cannot be context-sensitive entities as in the Corrected Fregean Theory. Hence, the view that demonstrations determine an object *in a context* must be abandoned, as explained by Kaplan (1989a, pp.528-529). Demonstrations determine objects on their own. It is then possible to include a sequence of objects determined by demonstrations - *demonstrata* - instead of the demonstrations themselves in the formal representation of context. This is the option usually taken by theorists.

Once demonstrations (or their objects) are moved to the context, sentences containing several occurrences of the same demonstrative pose a problem. Take 'This is bigger than this'. Assume that each occurrence of 'This' is associated with its own demonstration. If demonstrations appear in a sequence of parameters, rather than apposed to occurrences of demonstratives in logical form, how do we know which demonstration goes with which occurrence? One solution is to add subscripts to occurrences of demonstratives in logical form, so that e.g. 'This_1' is associated with the first element of the sequence of demonstrations of the context. It is the solution adopted by Kaplan for the Indexical Theory.[49]

With the addition of subscripts to syntactic occurrences of demonstratives, the Indexical Theory however faces a new problem. How can the meaning of demonstratives be formulated in this theory? Given the apparatus of subscripts, the most general statement we can make is that a demonstrative *d* with a subscript *n* refers to the object determined by the *n*-th demonstration of the sequence in the context. We cannot state any rule of the form: *'This' refers to the…* where 'This' is not subscripted (Braun 1996, pp.148-151). Formally, we could regard the meaning of demonstratives as a function from subscripts (indices) to characters. But for one thing this seems a convoluted way of capturing the simple idea that

[47] This view of the meaning of demonstratives is defended in (Braun 1996). The author however does not place demonstrations in the logical form of demonstrative sentences but rather in a sequence distinct from the list of parameters that represents context. The context itself contains a sequence of *demonstrata* (of objects determined by demonstrations). Braun's formal implementation is thus quite different from that of a corrected Corrected Fregean Theory.

[48] One might think that this is a bad idea because demonstrations are interpretation-dependent while speaker, time, and location are *interpretation-independent* (Brandom 2008, p.58). I discussed the interpretation-dependence of communicative intentions in Chapter 2.

[49] Another solution is to allow context to change mid-sentence, for example by changing the demonstration occupying the first place in the sequence (Braun 1996, pp.152-155).

demonstrations determine the reference of demonstratives. For another, this rule does not seem fit to act as a linguistic rule, i.e. something competent speakers of English know about demonstratives. In conclusion, the connection between Kaplan's Indexical Theory and desideratum (2) of classical metasemantics is not straightforward either.

So much for the relation between Kaplan's theory and desideratum (2). One desideratum of classical metasemantics remains unaccounted for. According to desideratum (1), a metasemantics for demonstratives must provide individually necessary and jointly sufficient conditions for an object to be their referent. I find it harder to trace back this desideratum to Kaplan's views. For one thing, even if demonstratives had characters, desideratum (1) would not follow. Take the following (hypothetical) reference-determining rule for 'This':

> 'This' refers to the object being pointed at if a pointing gesture is performed, and refers to the most salient object otherwise.

Putting aside issues about the formal representation of context, this rule provides a character: it associates the expression with an object given a context of utterance. However, the transposition of this rule as conditions on reference yields *individually sufficient but not necessary* conditions:

> If the speaker performs a pointing gesture and points at *o* in c, then *o* is the referent of 'This' in *c*.
> If the speaker performs no pointing gesture and *o* is the most salient object in *c*, then *o* is the referent of 'This' in *c*.

This list of sufficient conditions can be turned into a biconditional. But the right-hand side of the biconditional will be a disjunction:

> 'This' refers to *o* in *c* if and only if: either the speaker performs a pointing gesture and points at *o* in c, or the speaker performs no pointing gesture and *o* is the most salient object in *c*.

I can only offer a guess regarding the connection between Kaplan's views and desideratum (1) of classical metasemantics. Kaplan's theories of pure indexicals and demonstratives might both be formulated as: 'Expression *e* refers in *c* to the F' (where F is not disjunctive). Kaplan does present his theory of pure indexicals in this manner: *"In each of its utterances, 'I' refers to the person who utters it"* (1989a, p.520, emphasis mine). As for demonstratives, Kaplan's theory can in effect be formulated as: demonstrative *d* refers in *c* to the object of the accompanying demonstration in *c*. A theory of the form 'Expression *e* refers in *c* to the F' may be transposed as a biconditional that provides an individually necessary and sufficient condition:

> Expression *e* refers to *o* in *c* if and only if *o* is the F.

Gomez-Torrente's take on Kaplan's work supports this interpretation. Gomez-Torrente calls Kaplan's view of the linguistic meaning of indexicals and demonstratives *"descriptivist"* (e.g. Gomez-Torrente 2019, p.24). Gomez-Torrente further presents a metasemantics that only provides sufficient (but not necessary) conditions as an *"alternative to the Kaplanian descriptivist project"* (Gomez-Torrente 2019, p.47).

Now, despite his formulation of several reference-determining linguistic rules as 'Expression *e* refers to the F', nowhere does Kaplan claim that reference-determining linguistic rules should take this form. As we have just seen, an expression can have a character although this character is not stateable as 'Expression e refers to the F'. Hence, an expression can have a character although its metasemantics cannot be stated as a list of individually necessary and jointly sufficient conditions. Beyond character, an expression can have a meaning that encodes the determinants of its reference although the theory of these determinants cannot be stated as a list of individually necessary and jointly sufficient conditions.

As I see it then, desideratum (1) is not part of Kaplan's theory. Metasemanticists have assumed that the reference-determining rule associated with demonstratives must be stateable as 'Expression e refers to the F' because the rules put forward by Kaplan are stated in this way. In my opinion, this is what led metasemanticists to burden themselves with desideratum (1).

2. The case for going non-classical

Now that I have traced back the origin of the three desiderata that define classical metasemantics, I wish to present the case for going non-classical. The first reason to go non-classical is that no extant metasemantics seems to work *qua* classical metasemantics. As already explained, intentionalism struggles to satisfy desideratum (3) in general, and some variants of intentionalism are not plausible if desideratum (2) is endorsed. Let me continue with a quick tour of the various *non-intentionalist* classical metasemantics that have been argued to fail.

Some non-intentionalist metasemantics for demonstratives are, *qua* classical metasemantics, just as susceptible to the Carnap-Agnew case as intentionalism. Devitt thinks that the reference of a demonstrative is the object of *de re* thought the speaker expresses, which is itself determined by a causal relation (e.g. perception) holding between speaker and object (Devitt 2022). Devitt further argues that intentionalism is flawed in ways his own theory is not. This might be right, but his theory plausibly predicts that the speaker's demonstrative in the Carnap-Agnew case refers to Carnap's picture.[50] If Devitt's theory is advanced as a classical metasemantics, it falls foul of desideratum (3).

[50] See (Chapter 1, section 3b) for discussion of *de re* thought in the Carnap-Agnew case.

There is another family of non-intentionalist metasemantics that is likely to fail. According to *external cues* metasemantics, the publicly accessible cues given by the speaker about the object of her thought determine reference. Within this family of theories we find the *demonstration* theory according to which the referent of a demonstrative is the object determined by the speaker's ostensive gesture (McGinn 1981). The demonstration theory faces two problems that have been mentioned in Chapter 1: ostensive gestures are indeterminate, and there are non-ostensive cases of demonstrative reference. On a more inclusive view of external cues, external cues encompass not only ostensive gestures but also uttered sentences, and hence coincide with what I called the speaker's proximal act in Chapter 1. On the face of it, the speaker's proximal act determines just what the proximal intention attaching to this proximal act determines. If this is correct, an external cues metasemantics makes the same predictions as a proximal intentions (PI) intentionalism. Since PI intentionalism cannot secure the right predictions about reference, neither can *external cues* metasemantics. These metasemantics also flounder on desideratum (3).

Another oft-discussed theory is *salience* metasemantics. The referent of a demonstrative is taken to be the salient object - or the most salient object - in the context of utterance (Gauker 2008, Heck 2014, Mount 2008, Speaks 2017). According to one development of the notion of salience, to be salient is to have perceivable properties that tend to draw the attention of perceivers (Gauker 2008). Salience is then some kind of dispositional property grounded in perceivable properties. This conception of salience supports a metasemantics on which the referent is the *most* salient object in the context of utterance. On another conception of salience, to be salient is to be the object of the joint attention of speaker and hearer (Mount 2008). This conception of salience supports a metasemantics on which the referent is the salient object *tout court* in the context of utterance. Salience metasemantics has been argued to fail. Sometimes by uttering a demonstrative one says something about an object that is not the most salient object in the perceivable-dispositional sense (Gauker 2008, Heck 2014). This object might be as salient as some other object in the vicinity, or even less salient. In addition, one can say something about an object that is not *the* salient one in the joint attention sense. In various cases of demonstrative reference, several objects may be said to be jointly attended to by speaker and hearer although only one is the intuitive referent (Speaks 2017). Again, here is a family of metasemantics that fails to satisfy desideratum (3).

There are two metasemantic proposals on which the jury is still out. According to King (2013, 2014), it is necessary for demonstrative *d* to refer to object *o* that a non-actual audience with specific properties (e.g. being attentive and competent) takes *o* to be the intended referent.[51] Speaks (2016) and Nowak & Michaelson (2021) have cast doubt on this proposal, arguing that no characterisation of the non-actual audience yields the right predictions about reference in every case of demonstrative reference. So this metasemantic proposal potentially flounders on desideratum (3).

[51] Here I ignore the fact that King's metasemantics is also an intentionalist one: it is also necessary that *o* be targeted by an intention of the speaker for *o* to be the referent of a uttered demonstrative.

Finally, Stojníc et al. (2013, 2017) claim that the reference of demonstratives is not determined by the extra-linguistic context, but rather by a linguistic context that consists of a sequence of objects modifiable by linguistic update operations. In other words, a demonstrative refers to *o* if and only if *o* is at the top of an object-sequence which is in some sense linguistically determined. On the surface, extra-linguistic contextual features are indeed absent from this metasemantic proposal. However, several authors have argued that this is an illusion (García-Carpintero forthcoming, Nowak & Michaelson 2020). This point deserves a short explanation. Take the central case of the use of a demonstrative accompanied by an ostensive gesture - a 'deictic' use. According to Stojníc, ostensive gestures provoke a linguistic update that puts the ostended object at the top of the object-sequence. Here we seem to face the familiar problem that ostensive gestures are indeterminate, but Stojníc et al. deny this on the basis that ostensive gestures are partly individuated by the ostended object. This metaphysical stipulation raises the following question. What determines *which ostensive gesture* is performed in a context of use: is it intentions, salience, etc? The metasemantic question has merely been displaced, and with it the role of extra-linguistic context. The philosophical question at the heart of the metasemantics debate remains as open as ever.

This quick tour of non-intentionalist metasemantics is enough to show that they are not in a better position than intentionalist metasemantics. The failure of various metasemantic theories to satisfy some desiderata of classical metasemantics constitutes one reason to go non-classical. But there is another reason to consider going non-classical: satisfying both desiderata (2) and (3) seems particularly difficult. My first observation on this front is that some metasemantics are hard to square with desideratum (2), although they are ostensibly put forward as complying with it. King's *coordination metasemantics* is a good example (King 2013, 2014). Here is one formulation of King's metasemantics:

> A speaker S's use δ of a demonstrative expression in context c has o as its semantic value iff 1. S intends o to be the semantic value of δ in c; and 2. a competent, reasonable, attentive hearer H who knows the common ground of the conversation at the time S utters δ, and who has the properties attributed to the audience by the common ground at the time S utters δ would recognize that S intends o to be the semantic value of δ in c in the way S intends H to recognize her intention. (King 2013, p.290)

King further takes this metasemantics to be consistent with desideratum (2):

> I take the lexical meanings of demonstratives to require that a use of a demonstrative be supplemented by a speaker's intention [condition 1 above] that is recognizable by an ideal hearer [condition 2 above] (King 2014, p.229).

The problem is that King's metasemantics is not plausible as a linguistic rule known by speakers of the relevant language. I do not want to argue this on the basis of the sophistication of King's condition 2. My objection is rather that a linguistic rule including condition 2 does not make sense if a plausible picture of utterance-interpretation is endorsed. Here is this

picture. The overall task of an interpreter faced with the utterance of a demonstrative is to find out what or who the speaker is talking about. This is just the task of ascertaining *speaker reference*. Now, assuming that a linguistic rule associates the demonstrative with an object in a context of utterance, competent speakers of the relevant language automatically apply this rule whenever faced with the utterance of a demonstrative. The automatic application of the linguistic rule yields an object. If there is further evidence that the object yielded by this application - the semantic referent - is not the speaker's referent, the interpreter further tries to figure out which object distinct from the linguistically-determined object the speaker has in mind.

On this view of the interpretation of demonstratives, the application of an object-determining linguistic rule is then a default procedure for ascertaining speaker reference. Here is a nice formulation of this picture by Wettstein:

> Indeed the point of the institutionalized [linguistic] conventions I have been discussing is to facilitate communication concerning the item speakers have in mind and about which they intend to inform others, ask questions, and so on. In the most usual, everyday cases, moreover people refer to the items they intend, the items they have in mind [...], and in such cases the referent is just the intended referent. (Wettstein 1984, p.73)

If this picture is correct, object-determining linguistic rules are at the service of communication. Such rules reliably assist interpreters in figuring out what object the speaker refers to. As an aside, I should note that this picture holds only for object-determining linguistic rules that do not appeal to speaker reference. If a linguistic rule appeals to speaker reference - e.g. *'This' refers in a context to what the speaker refers to with her use of 'This' in this context* - applying the linguistic rule requires ascertaining speaker reference. The application of the rule then does not act as a default procedure for ascertaining speaker reference. Rather, the interpreter must *first* ascertain speaker reference in order to apply the rule correctly.

Here is the problem with condition 2 of King's metasemantics. Figuring out what some non-actual hearer with specific characteristics would take the speaker to intend does not facilitate the recognition of speaker reference. Why would the lexicon ask the interpreter to assess what *some non-actual hearer* would take the speaker to have in mind, when the task at hand is for *the interpreter herself* to figure out what the speaker has in mind? Not all object-determining rules raise this worry. For example, it would make sense for our lexicon to instruct us to associate an utterance of 'This' with *the object demonstrated by the speaker*: the demonstrated object is likely to be the one the speaker has in mind. It would also make sense for our lexicon to instruct us to associate an uttered demonstrative with *the most salient object in the context of utterance*: that object is likely to be the object the speaker has in mind. By contrast, King's second condition seems interpretively unhelpful. It saddles the interpreter with a task - i.e. figuring out what a non-actual hearer would take the speaker to intend - that gets in the way of the interpretive task.

Here is my diagnosis for this unpleasant feature. Although King does lip service to desideratum (2), he is primarily driven by desideratum (3) - i.e. predictive adequacy. Indeed, most of the work done to defend his preferred metasemantics centres on its handling of pre-theoretical judgements on possible cases (King 2013, 2014). Although several authors advance counterexamples (Speaks 2016, Nowak & Michaelson 2021), King's metasemantics makes intuitively correct predictions on a wide array of cases. So King's metasemantics does quite well on desideratum (3). In my opinion, the lesson to be drawn from King's endeavour is that satisfying both desiderata (2) and (3) is hard.

This conjecture can be further motivated by looking at one of Speaks' metasemantic proposals, which is supposed to surpass King's proposal with regard to desideratum (3) - i.e. predictive adequacy:

> The value of a use of a demonstrative d in a context c is o iff: [1] the speaker intends o to be the value of d in c [2] the speaker would be justified in believing that his audience will take o to be the object that the speaker intends to be the value. (Speaks 2016, p.330)

Again, I think that this metasemantics does not make sense as a linguistic rule given the role that such a rule is supposed to play in utterance-interpretation. The problematic condition of Speaks' metasemantics is condition [2]. Any rational, cooperative speaker must justifiably believe that the interpreter will take her to intend the object she does intend. And if the interpreter takes the speaker to be rational and cooperative, she (the interpreter) must take the intended object to be one the speaker justifiably believes she (the interpreter) will recognise as such. Finding out what object a rational and cooperative speaker intends is just finding out what object she justifiably believes her audience will recognise as such. Speaks' second condition is then redundant with regard to the interpretation of demonstratives. If object-determining linguistic rules are supposed to assist the interpreter in utterance-interpretation, then Speaks' linguistic rule does not fulfil this purpose. By contrast, Speaks' condition [2] is not redundant with regard to desideratum (3): the addition of this condition allows the theory to make correct predictions on a great number of cases. These predictions could not be made with condition [1] only. So here again is a metasemantics ostensibly designed to satisfy desideratum (3), and plausibly failing to satisfy desideratum (2).

Generalising from particular proposals by King and Speaks, it seems that achieving predictive adequacy - i.e. satisfying desideratum (3) - requires from a metasemantic theory to look beyond the speaker's intentions (and beyond the hearer's actual interpretation of the speaker's intentions). One must consider whether the speaker has done enough for her intentions to be recognised given the context of the utterance. But it is hard to see why the interpreter of a demonstrative should be concerned with the 'recognisability' of the speaker's intentions: the interpreter's task is just to recognise these intentions. Assessing whether the speaker's intentions are 'recognisable' looks like a completely different task, and one that does not help with the interpretive task. Assuming that object-determining linguistic rules play the role of a

default procedure in the interpretation of uttered demonstratives, they should not include conditions about the 'recognisability' of the speaker's intentions. If the foregoing is correct, desiderata (2) and (3) of classical metasemantics are in competition: a metasemantics that does well on (3) will not do well on (2).

Now that I have presented the case for going non-classical, I propose to explore different ways of going non-classical. The plan is to assess the consequences of the rejection of each of the three desiderata of classical metasemantics in turn.

3. Rejecting desideratum (1)

Here is desideratum (1) again:

> A metasemantics for demonstratives must provide individually necessary and jointly sufficient conditions for an object to be the referent of a demonstrative in context.

One metasemanticist who explicitly rejects desideratum (1) is Gomez-Torrente. In his own words, *"...there is just no realistic hope of finding a theory of jointly necessary and sufficient conditions for demonstrative reference…"* (Gomez-Torrente 2019, p.39). His preferred alternative is *"a theory of merely sufficient conditions for demonstrative reference"* (*ibid*, p.47). Gomez-Torrente proceeds to give a list of sufficient conditions that are intentionalist in spirit. For illustration, here are two of these merely sufficient conditions:

> If a speaker S forms the intention of using a demonstrative D to refer to an object as purportedly remembered by S, and the memory in question does correspond to a certain object o, then S's use of D will refer to o, if S forms no intention conflicting with that intention.

> If a speaker S forms the intention of using a demonstrative D to refer to an object as purportedly perceived by S, and the perception in question is in fact of a certain object o, then S's use of D will refer to o, if S forms no intention of using D to refer to a different object p as purportedly perceived by S, with the perception in question being in fact of p (adapted from Gomez-Torrente 2019, pp.51-52).

It is not clear to me why Gomez-Torrente distinguishes between several intentionalist conditions depending on the variety of mental reference involved. In any case, Gomez-Torrente is keen to present his metasemantics as a list of merely sufficient conditions.

Gauker also puts forward a metasemantics that does not provide a list of individually necessary and jointly sufficient conditions (Gauker 2008). This makes him a non-classical metasemanticist, although he does not explicitly discuss desideratum (1). Gauker provides a list of conditions (*"accessibility criteria"*) for an object to be the referent of a demonstrative.

These include being perceivable, having been previously referred to, and being pointed at by the speaker. Crucially, Gauker claims that none of these conditions is individually necessary or sufficient for an object to be the referent of a demonstrative. The referent of a demonstrative is the object that *"adequately and best satisfies"* the list of conditions (Gauker 2008, p.364). There is however no 'adequate and best satisfaction' algorithm. A verdict on reference is the result of an all-things considered judgement rather than that of an algorithmic procedure. Hence, there is no reference-determining linguistic *rule* associated with demonstratives. The fact that a demonstrative refers in context to the object that adequately and best satisfies a list of first-order conditions (being the most salient object, being pointed at, etc) is a matter of linguistic *convention,* but not of linguistic rule in the sense of an object-churning algorithm.

In the context of the present section, there is an interesting way of looking at Gauker's rejection of an 'adequate and best satisfaction' algorithm. This rejection may be regarded as the claim that there is no necessary or sufficient *higher-order condition*, in addition to there being no necessary or sufficient *first-order condition* for reference to obtain. Let me illustrate this point with examples. Say we have three first-order conditions C1-C3. None of these first-order conditions is individually necessary or sufficient. We can advance the following metasemantics on the basis of these first-order conditions: *o* is the referent of a demonstrative if and only if *o* satisfies at least two of C1-C3. This metasemantics provides a necessary and sufficient *higher-order condition* appealing to first-order conditions none of which is necessary or sufficient. By providing a higher-order condition, one provides a rule that takes as input a context of utterance and returns an object. Here is an example:

Step 1: return the object pointed at by the speaker if there is one. Otherwise, return #.
Step 2: return the most salient object if there is one. Otherwise, return #.
Step 3: return the last object referred to in the conversation if there is one. Otherwise, return #.
Step 4:

- If one and the same object is the output of at least two of the steps above, return that object.
- Otherwise, return #.

This rule takes as its input a context of utterance and returns an object. It then counts as a *character*, putting aside formal implementation in a model-theoretic semantics.

The moral of this incursion into higher-order conditions is double. First, once a distinction between first-order and second-order conditions is made, the space of possible metasemantics expands spectacularly. To give just one other example, one could design a metasemantics with neither-necessary-nor-sufficient first-order conditions and merely-sufficient second-order conditions. Building on the previous example: *o* is the referent of a demonstrative if and only if either *o* satisfies at least two of C1-C3, or *o* satisfies only one of C1-C3 and no other object does. The second lesson is that one does not need necessary and sufficient first-order conditions in order to get a rule that searches a context of utterance and returns an object. One

does not even need a necessary and sufficient *higher-order* condition. To this extent, desiderata (1) and (2) of classical metasemantics are independent: in principle, one can drop (1) while upholding (2).

Of course, some metasemantics that violate desideratum (1) struggle to satisfy (2). For example, the object-determining algorithm detailed on the previous page might seem too sophisticated to play the role of the *linguistic* rule associated with demonstratives by all competent speakers of the relevant language. It is nevertheless important to notice that this apparent over-sophistication is shared by some metasemantics that satisfy desideratum (1). King's metasemantics is a prime example:

> A speaker S's use δ of a demonstrative expression in context c has o as its semantic value iff 1. S intends o to be the semantic value of δ in c; and 2. a competent, reasonable, attentive hearer H who knows the common ground of the conversation at the time S utters δ, and who has the properties attributed to the audience by the common ground at the time S utters δ would recognize that S intends o to be the semantic value of δ in c in the way S intends H to recognize her intention. (King 2013, p.290)

If desideratum (1) - providing necessary and sufficient conditions - restricts the space of possible metasemantics without conferring an advantage to metasemantics that satisfy it, why have many metasemanticists assumed it? As I explained in the previous section, incidental features of Kaplan's theory of demonstratives make it seem as if the right answer to the question of the metasemantics of demonstratives has to take the form of individually necessary and jointly sufficient conditions. There might be a further explanation though. It is possible that philosophers have not been mindful enough of the nature of metasemantic theories. Individually necessary and jointly sufficient conditions are expected when one's theoretical goal is to answer a *real definition* question (e.g. What is demonstrative reference?) or a *conceptual analysis* question (e.g. What is the content of the concept REFERENCE?). But the question of the metasemantics of demonstratives is of a different kind. This question is: what makes it the case that a demonstrative such as 'This' or 'That' refers to a certain object when uttered by a speaker? It asks which facts determine the reference of a demonstrative. This question is a *grounding* question (Lewerentz & Marschall 2018). The answer to a grounding question need not take the form of a list of individually necessary and jointly sufficient conditions. Lewerentz and Marschall believe that some of the criticisms directed by Gauker at intentionalist metasemantics result from mistakenly taking metasemantics for real definitions.

As I see it, desideratum (1) can go. This frees up space for new metasemantics. However, I doubt that this move would save the metasemantic enterprise. Whether or not a metasemantics lists necessary and sufficient conditions, the tension between desiderata (2) and (3) remains: a metasemantics for demonstratives that does well on predictive adequacy is unlikely to provide a plausible object-determining linguistic rule, given the role such a rule is supposed to play in utterance-interpretation.

4. Rejecting desideratum (2)

The tension between desiderata (2) and (3) might tempt theorists into dropping one of the two. What are the consequences of dropping (2)?

> A metasemantics for demonstratives must double as the linguistic meaning of demonstratives: the determinants of reference are encoded in a linguistic rule mastered by speakers of the relevant language.

Suppose that certain facts - the metasemantics facts - determine reference facts. Suppose that our pre-theoretical say-judgements track reference facts. But further suppose that the linguistic meaning of demonstratives does not encode the metasemantic facts. Our language does not tell us what determines the reference of a demonstrative in a context of utterance. It might only tell us that e.g. 'This' should be used to talk about a relatively close (spatially or otherwise) entity.

One question raised by the possibility of a reference relation whose determinants are not linguistically encoded is that of the epistemic access to this relation, as mentioned in section 2.c. of the previous chapter. If competent speakers of the relevant language do not have access to reference through a linguistic meaning that encodes its determinants, how are they able to track reference with say-judgements? Do they enjoy a non-linguistically-mediated access to the determinants of reference? Or, as Åkerman suggests, do they have access to reference through facts other than its determinants? The theorist who chooses to drop desideratum (2) while upholding desideratum (3) must tell a plausible story about our epistemic access to reference-facts.

Another crucial question is that of the role played by the reference relation in communication if it is not the product of an automatic, default linguistic procedure for ascertaining speaker reference. On this front, there seems to be a problem for a non-linguistically-determined reference relation. Whatever the meaning of a demonstrative is, the interpretation of an utterance of it will start with the automatic deployment of this meaning by the interpreter. But unless this meaning instructs the interpreter to associate the expression with an object in a context of utterance, it is hard to see why the interpreter would do so apart from associating the expression with the speaker's referent. Take the linguistic meaning suggested in the previous paragraph: *'This' is used to talk about a relatively close (spatially or otherwise) entity*. When an interpreter faces an utterance of 'This', she automatically applies this rule, and thus tries to figure out which relatively close entity the speaker refers to with her utterance of 'This'. There seems to be no room in this process for a distinct relation between expression and object (distinct from speaker reference). If the determinants of the reference of demonstratives are not encoded in their linguistic meaning, the reference relation does not play a role in the explanation of communication with demonstratives.

To this it might be replied that interpreters of demonstratives who ascertain speaker reference pay attention to the speaker's ostensive gestures, to previous discourse, to perceptual salience, etc. All these features of the context of utterance can help the interpreter figure out what the speaker refers to. They are *broad* contextual facts, using terminology introduced in Chapter 2. Why couldn't one then say that the relation of interest is determined by one of these contextual facts (or a combination of them) and therefore that it plays a role in communication with demonstratives? Well, the reference relation would not play a *privileged* role in the interpretation of demonstratives. Whatever contextual facts determine it, the reference relation would not play a privileged role in the sense that other relations determined by other contextual facts would play a similar role, i.e. help the interpreter figure out speaker reference. Say that reference is determined by ostensive gestures: the referent is the object demonstrated by the speaker. There is also a relation determined by salience - the relation between the expression and the most salient object in the context. This relation plays a similar role to the demonstration-determined relation: it helps the interpreter figure out speaker reference. There is then nothing special about the relation whose determinants are investigated by metasemanticists when it comes to communication with demonstratives.

At this point, one might concede that the reference relation plays no role in the explanation of communication with demonstratives, and argue that it plays some further explanatory roles. Here is an illustration of this strategy. In Chapter 2 I introduced a metasemantics put forward by Neale and Schiffer (Neale & Schiffer 2020):

> x is the referent of d on an occasion of use o iff
> (1) Cx
> (1) the speaker refers to x with d on o.

C is a constraint on speaker reference encapsulated in the linguistic meaning of the expression. If *d* is 'She' for instance, C stands for *being female and distinct from the speaker*. If *d* is 'This', *C* stands for *being relatively close to the speaker*. Let us assume that this metasemantics is not linguistically encoded. And let us further assume that the linguistic meaning of demonstratives is just the constraint C: demonstrative *d* must be used to refer to something that satisfies C. The relation captured by this metasemantics is in effect a species of speaker reference. Semantic reference holds when a speaker uses an expression to speaker-refer to an object that satisfies the meaning of the expression. Capturing this relation allows us to make a distinction between meaning-compliant acts of speaker reference and meaning-violating acts of speaker reference.

This distinction is worth capturing. However, many metasemanticists would not want to call this relation 'semantic reference'. First, it is at best unclear that it is the relation tracked by pre-theoretical say-judgements: according to this metasemantics Agnew's picture is *not* the referent of 'That' in the Carnap-Agnew case for example. Secondly, most metasemanticists would not want semantic reference to be a species of speaker reference: for many theorists, the independence of semantic reference from speaker reference is something of a dogma. Finally, the relation characterised by a Neale-Schiffer metasemantics is not determined by

features of *the context of utterance* (unless speaker-reference is reclassified as a feature of the context of utterance). This constitutes in itself a significant departure from the kind of relation most metasemanticists have assumed to hold between demonstrative and object.

Now, there are other relations worth capturing than Neale and Schiffer's *meaning-compliant speaker reference*. In Chapter 1, I explained that Perry takes pre-theoretical say-judgements to (partially) track 'forensic content', a type of content that is not determined by the meaning of demonstratives (Perry 2009, p.191). Forensic content is the content the speaker is responsible for, in the sense that the audience is likely to take this content to be the one meant by the speaker. One is allowed to call 'reference' this forensic relation between expression and object which gives us a metasemantics that satisfies desideratum (3) but violates desideratum (2). However, it is important to realise that this forensic 'reference' relation is significantly different from the relation whose determinants are the topic of classical metasemantics. Granted, both relations are tracked by pre-theoretical say-judgements, and both are independent from speaker reference. But the 'classical' relation plays a privileged role in the explanation of communication with demonstratives, due to the linguistic encoding of its determinants. The 'non-classical' forensic relation does not play such a role. In the final chapter of this dissertation, I will argue that pre-theoretical say-judgements track something like Perry's forensic content.

Let me take stock. A theory of the determination of a relation that is tracked by pre-theoretical say-judgements but whose determinants are not linguistically encoded can be given. Indeed, the theory I will put forward in the final chapter of this dissertation is a theory of this kind. One may call such a theory a 'metasemantics', and call the corresponding relation 'semantic reference' if one likes. But one should keep in mind that this relation is not the relation that is assumed in classical metasemantics. The non-classical relation plays no role in the explanation of communication with demonstratives, because its determinants are not linguistically encoded.

5. Rejecting desideratum (3)

As a reminder, (3) says that:

> A metasemantics for demonstratives must make predictions about reference that match pre-theoretical say-judgements on cases (e.g. the judgement that the speaker said something about object *o*).

If one rejects desideratum (3), one thinks that a metasemantics does not stand or fall on the coincidence of its predictions with pre-theoretical say-judgements. There are several possible reasons for thinking this. One might think that pre-theoretical say-judgements simply do not track the relation of reference whose determinants are the topic of metasemantics. One might alternatively think that pre-theoretical say-judgements track this relation in general, but that there are local failures of the tracking mechanism. Finally, one might think that

pre-theoretical say-judgements track this relation, but not exclusively: they track this relation *and* other relations at the same time.

Let me start with the idea that say-judgements simply do not track the relation between expression and object whose determinants are the topic of metasemantics. The interaction of this idea with desideratum (2) is problematic. The situation may be presented as a dilemma: endorsing desideratum (2) has unpleasant consequences; but so does rejecting it. By endorsing desideratum (2), one accepts that the application of a linguistic rule yields an object associated with an uttered demonstrative. If one further assumes that the output of the linguistic rule is consciously available, then one cannot claim that say-judgements do not track reference without accepting that competent speakers are aware of a discrepancy between the object-output of the linguistic rule and the object of say-judgements in cases where they come apart. In other words, there should be cases in which we are aware that linguistic conventions pull us towards one object while say-judgements pull us towards another object. The problem is that there is no evidence of this phenomenon. We do register a discrepancy between what the speaker means and what the speaker says in cases such as the Carnap-Agnew case. But this discrepancy is not the discrepancy predicted by the view under consideration, unless what the speaker means is identified with the object-output of the linguistic rule; that is, unless speaker reference is identified with semantic reference.

One could try to block this line of reasoning by contending that the object-output of the linguistic rule is not consciously available to interpreters. Although not incoherent, this view should not appeal to metasemanticists. If the users of the linguistic rule have no conscious access to its output - i.e. the referent - then these users have no pre-theoretical grip on the relation determined by this linguistic rule. And without conscious access to this relation, there is no pre-theoretical anchor for the metasemantic enterprise. Åkerman suggests that there are further *theoretical* anchors for reference: theoretical roles that can be filled by a relation between expression and object (Åkerman 2015). As Åkerman acknowledges himself, different theoretical roles pull the theorist towards different metasemantics. The theorist might opt for an intentionalist metasemantics if she wants the relation of reference to capture the idea that speakers express their thoughts in making utterances. She might alternatively opt for an 'ideal audience' metasemantics *à la* King if she wants the relation of reference to capture the idea that speakers are responsible for what an audience with certain properties would take speakers to mean. However, if the theorist herself has no conscious access to the output of the object-determining linguistic rule associated with demonstratives - like every competent speaker of the language - she will never know whether the relation she has singled out on the basis of her theoretical priorities is the linguistically determined one. The metasemantic enterprise would then seem a futile one.

Here is the second horn of the dilemma for one who is tempted to reject desideratum (3) on the grounds that say-judgements do not track reference at all. She might choose to reject desideratum (2) rather than endorse it. According to the view under consideration, reference is neither determined by linguistic meaning nor tracked by say-judgements. As I argued in the previous section, forsaking desideratum (2) implies that a relation of reference distinct from

speaker-reference does not play a privileged role in communication with demonstratives. It is hard to see why one would want to theoretically single out a relation with the following profile: its determinants are not linguistically encoded; it plays no role in communication; and it is not tracked by say-judgements. What would this relation do for the theorist?

One could heed Åkerman's suggestion to focus on less obvious theoretical roles for a relation between demonstrative and object: mark the difference between meaning-compliant speaker reference and meaning-violating speaker reference as Neale and Schiffer do; capture the content the speaker is responsible for in some sense to be further specified… However, one and the same relation could not play all these further theoretical roles. The resulting view is then that there are several relations between demonstrative and object worth postulating: reference$_1$, reference$_2$, reference$_3$… Depending on theoretical priorities, the theorist can give pride of place to one or the other. This pragmatist, pluralist meta-theoretical picture could not be more at odds with the metasemantics for demonstratives debate as it has unfolded since Kaplan's pioneering contribution. This view is hardly distinguishable from the following *eliminativist* view: there is no such thing as the reference relation, although there are several relations between expression and object worth positing for various explanatory purposes. To be clear, I am not saying that a pragmatist, pluralist view about reference-relations is inherently problematic. But if one reaches such a view, nothing is left of the metasemantics project. This project assumes that there exists a relation between expression and object - reference - that has properties which make it central in a theory of meaning and language use: its determinants are linguistically encoded, it is tracked by pre-theoretical say-judgements, and it plays a crucial role in communication.

Let me now turn to the second way of rejecting desideratum (3). One can claim that say-judgements track reference in general, but that there are local failures of this mechanism. The tracking mechanism is then not fully reliable. I discussed the deployment of this strategy by intentionalists in Chapter 1. Here I will generalise the lessons learned there. The most straightforward way to pursue the idea that pre-theoretical judgements do not reliably track reference is that there are cases in which our usually accurate tracking of reference is inefficient.

This strategy is useful when a class of cases generates counterexamples to one's favoured metasemantics, but it risks being *ad hoc*. The first hurdle is to characterise the class (or the classes) of cases in which pre-theoretical reference-tracking fails. Indeed, it will not do to just list the cases that provide *prima facie* counterexamples to one's metasemantics and claim that reference-tracking fails precisely in *these* cases. This would constitute terminal ad hocery. Gomez-Torrente's deployment of the strategy, which I described in Chapter 1, meets this first bar. According to Gomez-Torrente, pre-theoretical reference-tracking is inefficient in cases of *conflicting referential intentions*.

The second hurdle is to provide empirical evidence that something unusual happens with our pre-theoretical say-judgements on cases in the defined class. Unless theorists agree that they have a common reaction to these cases, studies collecting the judgements of ordinary

speakers on several of these cases will prove absolutely necessary. Finally, the theorist who wants to claim that pre-theoretical reference-tracking fails in a certain class of cases must explain why it thus fails. In this regard, a distinction between two types of explanation can be drawn. The first type of explanation appeals to the object-determining linguistic rule associated with demonstratives, and thus requires the endorsement of desideratum (2) of classical metasemantics. This is the path taken by Gomez-Torrente. Gomez-Torrente contends that the linguistic rule associated with demonstratives is a function from a subset of the set of contexts of utterance to the set of objects. Very roughly, the rule associates an object with the expression in normal contexts of utterance only (in Gomez-Torrente's theory, a normal context is one in which the speaker does not have conflicting referential intentions). If certain cases of utterances of demonstratives are not covered by the linguistic rule associated with demonstratives, and if the output of this rule is generally available to consciousness, one should expect a strong phenomenological contrast between pre-theoretical judgements on cases covered by the rule and cases not covered by the rule. A *linguistic* explanation of local failures of reference-tracking thus comes with a heavy empirical burden. As I explained in section 2 of Chapter 1, Gomez-Torrente does not meet this empirical challenge.

According to the second type of explanation of local failures of reference-tracking, the linguistic rule associated with demonstratives covers every context of utterance. Assuming that the result of the application of this rule is consciously available, the idea is that a further mechanism interferes with the linguistic mechanism in the problematic class of cases, and thus 'mutes' the result of the linguistic mechanism. There are two ways this explanation can go: either one claims that the further mechanism is triggered only in the problematic class of cases, or one claims that it is triggered in every case but wreaks havoc in the problematic class of cases only.

Here is a possible application of the first strategy. An intentionalist metasemanticist can maintain that Carnap's picture is the referent in the Carnap-Agnew case, but that in this case and similar cases the speaker makes a mistake that is likely to result in miscommunication. This salient feature of the case detracts from reference and draws the attention of the observer on the misunderstanding that is likely to result from the speaker's mistake. This is why the observer judges that the speaker said something about Agnew's picture. Explaining away certain pre-theoretical judgements in this manner seems fine to me in principle: whether it is ad hoc depends on the detail of the explanation, and more specifically on how the detail resonates with the phenomenology of our judgements on cases. However, this type of explanation will always face the competition of a more straightforward explanation: reference is not determined in the way favoured by the theorist - here the speaker's intentions - and our pre-theoretical judgements on the problematic cases - here Carnap-Agnew cases - track reference as they do in every other case.

The second strategy consists in claiming that a general tracking mechanism other than the linguistic reference-tracking mechanism is always at play in pre-theoretical say-judgements, but that this further mechanism derails reference-tracking in the defined class of cases only. This is in effect what I presented as the third way to reject desideratum (3): claim that

pre-theoretical say-judgements track reference *and* other relations at the same time. I explained in section 2 of Chapter 1 that Perry deploys this strategy to defend intentionalism against the threat of Carnap-Agnew-style cases (Perry 2009). According to Perry, pre-theoretical say-judgements track the reference of demonstratives, which is intentionalist and linguistically encoded, but they also track 'forensic content'. Generalising from Perry's specific proposal, the strategy consists in claiming that pre-theoretical say-judgements track two relations at the same time, one of them being the linguistically determined relation of reference. Such double-tracking does not cause problems as long as the two relations converge on the same object, but creates confusion when the object *relata* diverge. Our problematic class of cases is thus analysed as one in which the results of this double-tracking diverge. If pre-theoretical judgements on the problematic cases go against the predictions of one's favoured metasemantics, one must add (as Perry does) that the non-reference relation tends to be dominant over the reference relation in our pre-theoretical judgements.

Some of the problems I listed for Perry's specific implementation of this strategy in Chapter 1 affect this strategy in general. First, if pre-theoretical say-judgements track two relations R1 and R2, on what basis should one claim that one relation is linguistically determined but not the other? This choice seems arbitrary. If there was a special phenomenology associated with applying an object-determining linguistic rule, the problematic cases would not be problematic: it would be obvious which of the two diverging objects is the linguistically determined referent. Secondly, the view under consideration is empirically indistinguishable from the view that pre-theoretical judgements track only the non-linguistically-determined relation. If pre-theoretical judgements tracked only the non-linguistically-determined relation in every case, we would get in every case *the same judgement* we would get if the double-tracking view was correct. Finally, the overall data can receive a much simpler explanation than double-tracking. Here is this explanation. The relation that the double-tracking view deems distinct from semantic reference *is* semantic reference after all, and it is the only relation tracked by say-judgements. In other words, why have a metasemantics that violates desideratum (3) plus a complicated double-tracking view, when one can have a single-tracking view that satisfies desideratum (3)?

It will prove useful to compare Perry's views with the views recently expressed by Nowak and Michaelson (Nowak & Michaelson 2022). Like Perry, these authors think that the pre-theoretical notion of what is said is subject to several pressures. According to the pre-theoretical notion, speakers have privileged access to what they say - the intentionalist pressure - and yet they can be mistaken about what they say - the forensic pressure (Nowak & Michaelson 2022, pp.137-138). There will be cases in which these two pressures pull us in opposite directions regarding what the speaker said. The Carnap-Agnew case is one of these cases. One way to understand this talk of 'several pressures' on the notion of what is said is that with pre-theoretical say-judgements we track several relations at the same time. Nowak and Michaelson describe some sort of pre-theoretical multi-tracking, just like Perry. Now, here is where Nowak and Michaelson diverge from Perry. According to Nowak and Michaelson, no relation among the several relations tracked by pre-theoretical say-judgements is the linguistically determined relation of reference. What the authors

propose instead is that we embrace a form of pluralism about reference. As they put it: *"there are multiple, though often coextensive, notions of semantic value [reference] and what is said available - each suited to different explanatory projects"* (Nowak & Michaelson 2022, p.151). The view on offer here is not that our language associates several object-determining rules with demonstratives, and thus instructs us to associate several (possibly distinct) objects with each uttered demonstrative. Nowak makes this explicit in another paper: *"lexicalized instructions for finding the referent of a demonstrative would be superfluous"* (Nowak 2019, p.200). According to Nowak, no object-determining linguistic rule is associated with demonstratives. Therefore, none of the relations tracked by our say-judgements is linguistically determined.

The multi-tracking view of Nowak and Michaelson thus comes with a pluralist, pragmatist meta-theoretical picture of the sort I discussed earlier in this section. On this picture, there are several theoretically interesting relations between demonstrative and object, none of which has its determinants encoded in the meaning of demonstratives. Nowak and Michaelson thus further reject desideratum (2) of classical metasemantics. As I have already explained, a pluralist, pragmatist view about reference may be regarded as an eliminativist view about reference: although there are several theoretically interesting relations between demonstrative and object, there is no such thing as *the* relation of reference.

In summary, different ways of rejecting desideratum (3) lead to different results. First, claiming that pre-theoretical say-judgements do not track the reference relation at all leads to a pragmatist, pluralist picture at odds with the metasemantics for demonstratives project. Secondly, claiming that pre-theoretical say-judgements track reference but not reliably so risks being ad hoc without a detailed explanation of how reference-tracking works. Finally, claiming that pre-theoretical say-judgements track several relations at the same time leads again to a pragmatist, pluralist picture.

Conclusion

Here are the three desiderata that define classical metasemantics. A metasemantics for demonstratives must:

(1) Provide individually necessary and jointly sufficient conditions for an object to be the referent of a demonstrative in context.
(2) Double as the linguistic meaning of demonstratives: the determinants of reference are encoded in a linguistic rule mastered by speakers of the relevant language.
(3) Make predictions about reference that match pre-theoretical say-judgements on cases (e.g. the judgement that the speaker said something about object *o*).

Desiderata (1) and (2) can be traced back to Kaplan's work (section 1). It is not clear that all three desiderata can be jointly satisfied: many extant metasemantics flounder on at least one of them; and there are reasons to think that desiderata (2)-(3) cannot be jointly satisfied

(section 2). Desideratum (1) seems to me an unnecessary theoretical harness for metasemanticists (section 3). Desideratum (2) can be dropped, but a theory constrained only by (3) assumes a relation of reference quite different from the one assumed by classical metasemantics (section 4). Desideratum (3) can be rejected in different ways, which come with various difficulties. Most interestingly, several ways of rejecting (3) lead to a pragmatist, pluralist view about reference which is at odds with the spirit of the metasemantic enterprise (section 5).

Chapter 5: Publicity, productivity, and semanticity

In section 2 of the previous chapter I argued that the project of finding the true metasemantics for demonstratives is likely to remain fruitless as long as the reference relation is assumed both (i) to be tracked by pre-theoretical say-judgements and (ii) to have its determinants encoded in a linguistic rule. While the metasemantics for demonstratives project as classically conceived simply assumes this, there are considerations in favour of the view that pre-theoretical say-judgements track a relation whose determinants are linguistically encoded. In this chapter I give a fair hearing to these considerations, before arguing that they ultimately do not support the view in question.

1. Publicity and semanticity

In this section I want to argue that the content of pre-theoretical say-judgements is neither determined by the speaker's intentions nor by her audience's actual interpretation. I call this property - independence from both speaker intentions and audience interpretation - *publicity*. The thesis I want to defend is then that the content of pre-theoretical say-judgements is *public*. To be in a position to defend it, I will first defend a weaker thesis: the content of pre-theoretical say-judgements can come apart from both the content intended by the speaker and the content retrieved by her audience. I will sometimes use the phrases 'what is said' and 'what the speaker says' as alternatives to the longer 'the content of pre-theoretical say-judgements'.[52]

a. Double divergence of what is said

That people sometimes say something they do not mean is a pre-theoretical platitude. Philosophers may build theoretical notions of *saying* and *meaning* such that one cannot say something without meaning it, but this is a case of revisionary theorising (Unnsteinsson 2019). Here I want to do justice to the pre-theoretical platitude. If speaker-meaning is analysed in terms of communicative intentions, the pre-theoretical platitude can be reformulated as: what a speaker says sometimes comes apart from what she communicatively intends. In Chapter 1 and Chapter 3, we saw that some authors attempt to make what is said dependent on intentions other than the usual communicative intentions, partly in response to the observation that what a speaker means can come apart from what she says. And I argued that these attempts are misguided. Here I allow myself to ignore these attempts, taking the possible divergence between what the speaker means and what she says as evidence that what the speaker says is not determined by the speaker's intentions.

[52] 'What is said' has sometimes been used as a term of art, notably by Kent Bach (2001). But most authors have used it as I use it, to pick out the 'intuitive truth-conditions' of utterances of declarative sentences (Recanati 2004, Stanley 2007).

Of course, a speaker can communicatively intend several contents. Speakers often mean what they say *and* something else beyond what they say. Once this commonplace is taken into account, the thesis that what is said sometimes comes apart from what is meant may seem trivial or confused. To correct this, one must distinguish between different kinds of communicatively intended content - i.e. different kinds of speaker-meaning. Various authors make a distinction between primary and secondary speaker-meant content (Bach & Harnish 1979, Camp 2012, Carston 2002, Recanati 2004). Very roughly, a secondary speaker-meant content is one that the speaker expects her audience to retrieve on the basis of retrieving another speaker-meant content. A secondary speaker-meant content is thus dependent on another speaker-meant content, while a primary speaker-meant content is not dependent in this way. An example will help. You ask me whether I want to watch Ang Lee's *Sense and Sensibility*. I reply: 'I love period dramas!'. I mean that I love period dramas, but of course I also mean that I want to watch *Sense and Sensibility*. I expect you to recognise that I mean the latter content on the basis of recognising that I mean the former content. The former content is primary; the latter content is secondary. The possible divergence I have in mind is then one between what the speaker says and what the speaker *primarily* means with her utterance. I will sometimes drop the qualification, so one should keep in mind that the speaker-meant content in question is the primary one.

There are several ways for what a speaker says to come apart from what she primarily means. Sometimes a speaker *intentionally* says something she does not primarily mean. Intentional divergence of this sort is illustrated by sarcastic utterances, as when I utter 'He's a fine friend indeed' in response to a tale of back-stabbing.[53] *Unintentional* divergence between what the speaker says and what she primarily means is a different beast, which we have already encountered in the guise of counterexamples to intentionalism. In the Carnap-Agnew case for example, it seems that the speaker primarily means something about Carnap's picture but says something about Agnew's picture. This is unintentional.

Cases of unintentional divergence between what is said and what is primarily meant may actually be divided into two categories. A speaker can say something other than what she means *by accident* or *by mistake*. I borrow this distinction from (Austin 1957). Austin makes this distinction in the broader realm of action: one can do something one does not intend by accident or by mistake. For example, I may unintentionally stroke donkey A. This could be e.g. because I intend to stroke donkey B but mistake donkey A for donkey B. In this case I stroke donkey A *by mistake*. Alternatively, I may unintentionally stroke donkey A because I have poor motor control and my hand ends up on donkey A although I intended to stroke donkey B. In this latter case, I stroke donkey A *by accident*. Very roughly, the distinction between mistake and accident tracks whether the divergence between the intended act and the act actually performed is explainable by a false belief in the agent's plan or by a failed execution of the plan, respectively.

[53] This respects the pre-theoretical platitude that sarcastic speakers say something they do not mean. See (Camp 2011) for an introduction to theories of sarcasm.

Austin's distinction can be transposed to the unintentional divergence between what a speaker communicatively intends and what she says. Take the Carnap-Agnew case. In the original version of the case, divergence occurs because the speaker falsely believes that the picture behind her is Carnap's picture. She says something she does not mean *by mistake*. Now, there is a variant of this case in which divergence occurs because the speaker has poor motor control and points behind her when she really wanted to point to her left, where Carnap's picture is actually located. In this variant, the speaker says something she does not mean *by accident*.

An orthogonal distinction can be drawn that pertains to the aspect of the utterance affected by accidents or mistakes. In both variants of the Carnap-Agnew case just mentioned, the mistake/accident affects the ostensive gesture of the speaker. These cases can be contrasted with the following case. A speaker makes a gesture in the direction of a man and a woman sitting at a restaurant table, uttering 'He seems nice' while having the woman in mind. Intuitively, the speaker said that *the man* is nice. Whether the speaker did this by mistake or by accident (I am assuming that the misgendering at hand is unintentional), it is the linguistic part of the utterance that is affected. Accidents affecting the linguistic part of an utterance are pre-theoretically known as 'slips of the tongue'.

A final distinction can be drawn among mistakes affecting the linguistic part of an utterance. *Linguistic* mistakes are false beliefs about properties of the linguistic expressions one utters. As explained in (Reiland 2023), some philosophers have denied that there are such things as linguistic mistakes. Again, this is a piece of revisionary theorising. Pre-theoretically, it is obvious that sometimes speakers are mistaken about the meaning of an expression, and that this results in saying things they do not mean. Here is a simple example. I believe that 'evasive' means *elusive* in English. I utter 'Answers remain evasive', meaning that answers remain elusive. I have said by mistake that answers remain evasive. In contrast with linguistic mistakes, *non-linguistic* mistakes are false beliefs about non-linguistic facts. Suppose I am watching the crowd of a restaurant from the bar. I whisper to my friend: 'The man in the green shirt is handsome'. The man I have in mind is actually wearing a blue shirt, and he happens to be sitting not too far from a handsome man in a green shirt, whom I failed to notice. I meant something about the man in the blue shirt, but intuitively I ended up saying something about the man in the green shirt. My mistake affects the linguistic part of my utterance, but it does not consist in a false belief about meaning. It rather consists in a false belief about the colour of the shirt worn by the man I have in mind.[54]

[54] In the philosophical literature, linguistic mistakes and slips of the tongue are usually grouped together under the label of 'malapropisms' (Davidson 1986, Predelli 2010, Reimer 2004, Unnsteinsson 2017). This theoretical grouping is awkward. In one case the speaker says something she does not mean by mistake, while in the other she does so by accident. Now, if one wants to have a theoretical category that encompasses both mistakes and accidents affecting the linguistic side of an utterance, this category should further include *non-linguistic* mistakes (e.g. uttering 'The man in the green shirt is handsome' having in mind a man who is actually wearing a blue shirt).

I have reviewed the various ways in which a speaker can unintentionally say something other than what she primarily means. These cases show that meaning that *p* is neither sufficient nor necessary for saying that *p*. I shall be briefer about the divergence between what the speaker says and what her audience takes her to mean. It is yet another pre-theoretical platitude that the audience sometimes understands something else than what the speaker said. Putting aside cases in which the audience correctly recognises that the speaker intentionally says something she does not mean (e.g. sarcasm), divergence of this sort typically occurs when the audience makes factual or linguistic mistakes similar to those made by speakers. The speaker says that *p* and means it, but due to an interpretive mistake the audience takes the speaker to mean that *q*. Divergence between what the speaker says and the audience's interpretation also occurs when the audience sees through an accident or mistake affecting the speaker's utterance. Imagine a speaker accidentally uttering 'Answers remain evasive', meaning that the answers remain elusive. And suppose that the audience understands that the speaker's tongue slipped, and that she meant that *answers remain elusive*. The speaker did say that *answers remain evasive*, but her audience recognises that she meant that *the answers remain elusive*. These several possibilities of divergence between what the speaker says and what her audience takes her to mean show that being taken by one's audience to mean that *p* is neither necessary nor sufficient for saying that *p*.

I have established that there are cases in which the speaker means that *p* but says that *q*, and that there are cases in which the audience takes the speaker to mean that *p* but the speaker says that *q*. Now, what is said sometimes comes apart from *both* what the speaker means and what the audience takes her to mean. There are thus cases of *double divergence*. We need not go far to find cases of double divergence. Take a variant of the Carnap-Agnew case in which the audience knows about the speaker's Carnap obsession, about the picture of Carnap usually hanging on the wall, etc. When the speaker points at the picture of Agnew, which is now hanging where the picture of Carnap usually hangs, and utters e.g. 'That's my favourite picture', this audience might see through the speaker's mistake and recognise that the speaker is talking about Carnap's picture. The speaker means something about Carnap's picture, and the audience takes the speaker to mean something about Carnap's picture. Communication succeeds. But what does the speaker say? My own judgement is that the speaker still says something about Agnew's picture, although she successfully communicates to her audience her thought about Carnap's picture.

Several authors seem conflicted about this case of double divergence. As I mentioned in section 2 of Chapter 1, Perry makes the interesting claim that our pre-theoretical judgements exhibit some sensitivity to what the audience is likely to understand. According to Perry, when the audience is likely to see through the speaker's mistake we are more ready to judge that the speaker said what she meant, despite her mistake. In the relevant variant of the Carnap-Agnew case, this would be a judgement that the speaker said something about Carnap's picture despite pointing at Agnew's picture. However, Perry concedes that the inclination to judge that the speaker said something about Agnew's picture is stronger. King also thinks that pre-theoretical judgements are affected by whether the audience is likely to see through the speaker's mistake (King 2013, pp.296-302). However, his preferred

metasemantics in the cited article makes the prediction that Agnew's picture is the referent, whether or not the audience sees through the speaker's mistake.

This alleged variability in pre-theoretical judgements could be empirically tested, but even the authors who suggest it do so timidly. In addition, there are cases of double divergence which do not involve the audience 'seeing through' a mistake or accident affecting the speaker's utterance. Sometimes both speaker and hearer make a linguistic mistake, for example. The speaker might utter 'Answers remain evasive' meaning that answers remain elusive, while the audience also commits a linguistic mistake, thinking that 'evasive' means *abrasive*. There might even be cases in which speaker and hearer make the same mistake. On this front, I have been told by native English speakers that non-native speakers (a class to which I belong) tend to be under common linguistic misapprehensions. This may create cases of double divergence. Finally, both speaker and audience could make *non-linguistic* mistakes. There is a variant of the Carnap-Agnew case in which the speaker points at Agnew's picture, and a distracted audience takes her to pointing at *Quine's picture*, which hangs some distance away from Agnew's picture on the wall behind the speaker. The speaker means something about Carnap's picture, her audience takes her to mean something about Quine's picture. What does the speaker *say*? My own judgement is still that the speaker says something about Agnew's picture.

I admit that my intuition loses some of its force in this last case. Whereas a double linguistic mistake leaves no doubt as to what is said (or so it seems to me), a double factual mistake affecting the reference of a demonstrative seems to render judgements on what is said difficult. Here is a preliminary diagnosis of this phenomenon. When the speaker makes a factual mistake but the audience does not, we side with the interpretation of the audience. When the audience makes a factual mistake but the speaker does not, we side with the communicative intention of the speaker. These cases are easy to settle as external observers. But when both speaker and hearer make factual mistakes, there is no one left to side with. And we seemingly cannot fall back on the content determined by or consistent with conventional meaning, in contrast with cases of double linguistic mistake. To judge what the speaker says in cases of *double non-linguistic mistake*, it seems that we must think about what a non-actual audience who makes no mistake (unlike the actual audience) would take the speaker to mean.

b. From double divergence to publicity

Assuming now that there are cases of double divergence of what is said - from what the speaker means and what her audience takes her to mean - does this establish that what is said is *public*? It is first important to notice that the non-necessity and non-sufficiency of each of the speaker's intention and the audience's actual interpretation does not entail that what is said is independent from these two factors. There are indeed possible views according to which some facts are neither necessary nor sufficient for another fact to obtain, and yet the former facts are determinants of the latter fact. The idea here is that the former facts act as

'part-time determinants': they are determinative in some cases but not others. In the case of what is said, we observed that sometimes what a speaker says is not what she means when her utterance is affected by a mistake or accident. We also observed that what the speaker says is not what her audience takes her to mean when the audience's interpretation involves a mistake. The following view, which I present for the sake of illustration, is consistent with these observations:

> S says that *p* if and only if: (no mistake/accident in the speaker's utterance and S means that *p*); or (no mistake in the audience's interpretation and the audience takes the speaker to mean that *p*).

On this view, the speaker's intention is determinative when no mistake or accident affects her utterance, while the audience's interpretation is determinative when no mistake affects interpretation. The first point of putting forward this toy theory is to show that meaning that *p* and being taken to mean that *p* could be part-time determinants of saying that *p*, and hence that the non-necessity and non-sufficiency of these conditions do not quite entail that they are not determinants of what is said. We could have a *variable-determinants* theory of determination instead of an orthodox *fixed-determinants* theory.

The second point of presenting the variable-determinants theory above is to show that it is inconsistent with cases of *double divergence*. Indeed, the toy theory above does not allow a speaker to say that *p* when (i) she does not mean that *p* and (ii) her audience does not take her to mean that *p*. Should we then conclude from cases of double divergence that what is said is independent from the speaker's intention and the audience's interpretation - is *public*? I think we should, but I must also note that publicity does not quite follow from there being cases of double divergence. In principle, a more sophisticated variable-determinants theory could be devised in which a *third* variable determinant springs into action when neither the speaker's intention nor the audience's interpretation are determinative. The determination job is done in alternation by three kinds of facts rather than two. Facts of the third kind intervene in cases of double divergence.

One worry about a three-variable-determinants theory is that the first two determinants will turn out to be redundant. Suppose that the third determinant is the interpretation of a non-actual audience. If one holds that the interpretation of a non-actual audience determines what the speaker says in cases of double divergence, why not hold that it *always* determines what the speaker says? A *three-variable-determinants theory* risks collapsing into a *single-fixed-determinant* theory in which meaning that *p* and being taken to mean that *p* are redundant. The third determinant does all the determination work, all the time. Unless a three-variable-determinants theory is provided in which the speaker's intention and the audience's interpretation are not redundant, one should conclude from the double divergence

of what is said from what the speaker means and what the audience understands that what is said is independent from both - that what is said is *public*.[55]

c. From publicity to semanticity?

The publicity of the content of pre-theoretical say-judgements is consistent with, and easily explainable by the linguistically-determined nature of this content - i.e. its *semanticity*. In the introduction to this dissertation I gave a more precise characterisation of the property of semanticity. The content in context of a target expression is semantic if and only if it is fully determined by (i) the meaning of the expressions contained in the target expression; (ii) rules of meaning-composition; and (iii) the aspects of context of utterance invoked by the meaning of the contained expressions.[56] This characterisation allows for more or less direct forms of semanticity - i.e. of linguistic determination. Considering a simple expression, its content in context may simply be identical with its meaning, or it may alternatively be distinct from meaning but determined by it plus the aspects of the context of utterance invoked by meaning. This latter, indirect form of linguistic determination is exemplified by the reference of *indexicals* such as 'I' and 'Today'.

It should be noted that indirect semanticity is still a high bar to pass. Suppose that the conventional meaning of an expression merely restricts its content in context to a certain class of objects or properties. For example, the conventional meaning of 'He' may be regarded as a restriction of its content in context to *male beings distinct from the speaker*. Given this conventional meaning, the object about which the speaker says something with 'He' does not count as *semantic*. For the meaning of 'He' does not determine its content in context, despite constraining it. Now, one may regard the difference between constraining content and determining content as one of degree rather than kind. Content-constraining meanings restrict content to whatever bears a multiply instantiable property - e.g. being male and distinct from the speaker - while content-determining meanings restrict content to whatever bears a uniquely instantiable property - e.g. being the utterer of a given token of 'I'. Still, meanings

---

[55] Beyond three-determinants theories, one finds *particularist* theories. According to particularist theories, in every possible case the fact to be explained - here that the speaker says that *p* - has determinants, but it is not possible to make true generalisations about the determination of facts of the kind to be explained (Dancy 2017). A related class of theories is *pro tanto* factors theories, according to which there are strictly speaking no determinants but only favourable conditions. We encountered a *pro tanto* factors theory of the reference of demonstratives in the previous chapter: according to Gauker (2008), various facts are relevant to what a demonstrative refers to, but none is determinative. Ascertaining reference is an all-things-considered judgement that takes into account these various factors. A similar view could be advanced about the content of pre-theoretical say-judgements. Particularist and *pro tanto* views constitute significant theoretical retreats. By endorsing such views, one acknowledges that neither a fixed-determinants nor a systematic variable-determinants theory can be given.

[56] In the introduction I presented a more sophisticated definition of *semanticity* that allows determination by unpronounced constituents in logical form. It is not necessary to take into account this refinement here.

of the latter type are designed to pick out a unique entity, while meanings of the former type are not so designed.

According to a semantic view of the content of pre-theoretical say-judgements, declarative sentences linguistically determine a proposition when uttered in context, and pre-theoretical say-judgements on these sentences track this linguistically-determined proposition (Stanley 2007). This view can easily account for the publicity of what is said. Speaker and audience may be mistaken about the meaning of an uttered expression, or about the context of utterance. When the speaker is mistaken, the linguistically determined proposition may come apart from the proposition she takes to be the linguistically determined one. Assuming that the speaker primarily means the proposition she takes to be linguistically determined, what she says comes apart from what she primarily means. On the audience's side, when the audience is mistaken about meaning or context, the linguistically determined proposition may come apart from the proposition the audience takes to be so determined. If in addition the audience takes the speaker to primarily mean the proposition she (the audience) thinks is linguistically determined, what the audience takes the speaker to mean comes apart from what the speaker says. Hence, the semanticity of what is said can easily account for its publicity.

I should further note that the semanticity of what is said does not entail its publicity. A semanticist about what is said may independently hold that the content of context-sensitive expressions is indirectly determined by linguistic rules that invoke the communicative intentions of the speaker - i.e. what the speaker means.[57] On this view, what is said is semantic yet partly determined by what the speaker means. Semanticism about what is said is merely *consistent with* the publicity of what is said, and semanticity can easily explain publicity.

The capacity to account for publicity favours semanticism over some rival views of the determinants of what is said. According to a view we may call *pragmaticism,* what the speaker says with the utterance of a sentence outstrips what is linguistically determined, and it is a speaker-meant proposition (Carston 2002, Recanati 2004). What a speaker says is thus partly determined by what the speaker means. This view is plainly inconsistent with publicity, since it does not allow what is said to come apart from what the speaker (primarily) means. This is enough to grant semanticism an advantage over pragmaticism.

So semanticity can easily explain publicity. But is the publicity of what is said *best* explained by the semanticity of what is said? Without further ado: I think it isn't. The reason is simple. Pre-theoretical judgements on the content of speech acts other than *saying* also exhibit publicity, and some of these contents are uncontroversially not linguistically determined. Assuming that a unified explanation of the publicity of various speech act contents can be

[57] Stanley (2007) is a semanticist about what is said who seems sympathetic to the idea that the content of context-sensitive expressions is determined by linguistic rules appealing to the communicative intentions of speakers, although he officially stays neutral about the features of contexts invoked by these rules.

given, the publicity of what is said is not best explained by its semanticity. I devote the rest of this subsection to fleshing out this argument.

At least some of the things we judge speakers to imply (suggest, insinuate) are not linguistically determined. This much is uncontroversial. Grice's category of *conversational implicatures* is supposed to capture contents that are conveyed by the utterance of a sentence but are neither (i) determined by the linguistic meaning of this sentence nor (ii) identical with what the speaker said (Grice 1989). Now, the *publicity* of such contents is a more controversial issue. To tackle it, it will prove useful to separate two questions: whether Grice's conversational implicatures are supposed to be public; and whether what we pre-theoretically judge speakers to imply (suggest, insinuate) is public.

In a string of publications, Jennifer Saul argues that Grice's conversational implicatures are not fully determined by what speakers mean beyond what they say (Saul 2001; 2002; 2010). In her own words: "*What Grice's theory gives us [...] is the idea that what is implicated is not wholly up to the speaker*" (Saul 2001, p.633).[58] According to Saul, meaning a proposition *p* is necessary but not sufficient for *implicating* it in Grice's sense. Another necessary condition is (roughly) that the speaker's audience would take the speaker to mean that *p* by competently reasoning from conversational principles. We get the following determinants for *conversationally implicating that p*:

> Speaker S implicates that *p* if and only if:
> (i) S (secondarily) means that *p.*
> (ii) The speaker's audience would take S to mean that *p* by competently reasoning from conversational principles.

What is conversationally implicated is partly determined by what the speaker (secondarily) means, and partly determined by what a competent interpreter would take the speaker to mean. Saul calls the property of being partly determined by what a competent interpreter would take the speaker to mean 'normativity'. It is crucial to note that normativity does not amount to publicity. The publicity of a content *p* that the speaker Xs (says, implies, or whatever) entails that a speaker can X that *p* while meaning that *q*. But the mere normativity of this content does not allow this. At most, a speaker may fail to X that *p* when she means it.

Saul is not definitive about her interpretation of Grice's theory. Although she likes the 'normativity' interpretation, she seems tempted by the distinct interpretation that conversational implicatures are not even partly determined by speaker meaning: "*It is genuinely unclear whether Grice took being meant by the speaker to be a necessary condition for being a conversational implicature'* (Saul 2001, p.635). Given Saul's further insistence that the audience's actual interpretation does not play a part in determining what is

[58] Saul's (2001) is a review of (Davis 1998). Davis takes Grice's theory to be inconsistent because conversational implicatures cannot both be entirely a matter of speaker meaning and satisfy other conditions laid out by Grice. In essence, Saul's response is that Grice never intended conversational implicatures to be entirely a matter of speaker meaning.

conversationally implicated (Saul 2002, pp.242-243), it looks like Saul is tempted by an interpretation of Grice's theory according to which conversationally implicated content is *public*.

Though Saul's contribution mainly focuses on Gricean exegesis, it is highly valuable for my purposes in this chapter. Saul attacks the view that what is conversationally implicated is fully determined by speaker meaning, and she does so partly by checking the predictions of this view against pre-theoretical judgements on cases. Here is an example of this strategy, bearing in mind that Saul takes (Davis 1998) to defend a speaker meaning interpretation of Grice's theory:

> Albert tells Bettina that he feels sick. Bettina replies with 'I'll go find an aspirin'. Unbeknownst to Albert, what Bettina means by this is that there are aliens nearby, probably causing Albert's illness, and he should flee. (Bettina likes to protect herself from the aliens by holding an aspirin in front of her, and so she thinks that mentioning that she's going to find an aspirin is a good way to warn Albert. Albert knows nothing of Bettina's beliefs about aliens.) There is, it seems to me, little temptation to say that Bettina has implicated that Albert should flee. But on Davis' account this is just what she has done (Saul 2001, p.633, emphasis mine)

Saul's strategy in this passage is clear. She presents a case and appeals to our pre-theoretical judgement on it. This is unfortunately obscured by the fact that Saul writes as if the relevant pre-theoretical judgements involve the theoretical notion of *implicating*. Saul recognises herself that this is problematic but does not seem clear about her own methodology (Saul 2002, p.241). What is really going on, in my opinion, is that Saul appeals to a pre-theoretical judgement about what Bettina implies (suggests, hints…). We do make such judgements all the time, and Grice himself takes the notion of what people imply (suggest, hint…) as the pre-theoretical stepping stone to his notion of *conversationally implicating*.

Turning to the *Albert and Bettina* case, a speaker-meaning view of conversational implicature predicts that Bettina conversationally implicates that Albert should flee. But this does not match our pre-theoretical judgement on the case: Bettina does not imply this. I am not sure how Saul intends to argue from here. She might continue that charitability requires that we do not attribute to Grice a theory whose predictions clash so egregiously with pre-theoretical judgements. Or she might continue that Grice demonstrably intends his theory to be in harmony with pre-theoretical judgements. Or she might continue that *we* contemporary theorists should not want a theory whose predictions clash with pre-theoretical judgements. Whatever Saul wants to do with cases of this sort, I think that they can be used to show that meaning that *p* is neither sufficient nor necessary for implying that *p*. Let me slightly alter the dialogue of *Albert and Bettina* while keeping Bettina's beliefs about aliens and aspirins intact:

Albert: I've got a headache.
Bettina: There's aspirin in the bathroom cabinet.

Bettina means that Albert should flee. But intuitively that is not what she implied. What she implied is indeterminate, but the following propositions are intuitive candidates: that Albert should ingest an aspirin, that ingesting an aspirin will soothe his headache… Bettina thus implies one or several propositions, but she does not mean any of these propositions. This case seems to show that meaning that *p* is neither sufficient nor necessary to imply that *p*.

It is important to note that *Albert and Bettina* is under-described in an important way. Saul tells us that Albert knows nothing of Bettina's beliefs about aliens and aspirins. But does Bettina know that Albert knows nothing of her beliefs about aliens and aspirins? Answering this question one way or the other makes a difference. If Bettina knows that Albert does not know, then she is being irrational. But this irrationality has nothing to do with outlandish beliefs about aliens and aspirins! This irrationality could actually take two forms. First, Bettina might reasonably expect that Albert will *not* retrieve the content she wants to communicate. But then it is not clear that Bettina genuinely *means* that Albert should flee, and not clear why she produces her utterance. Several theorists have stressed that one does not have a genuine communicative intention unless one believes that one's audience is able to recognise it (Bach 1994a, Neale 2016). This is just a local application of a general constraint on the formation of intentions. As Grice put it, '*it is in general true that one cannot have intentions to achieve results which one sees no chance of achieving'* (Grice 1989, p.98). The second form that Bettina's irrationality could take is the following. She might expect Albert to recognise her communicative intention. Expecting this from Albert while knowing that he knows nothing of her beliefs about aliens and aspirins is irrational. Assuming that this second sort of irrationality is consistent with genuine speaker meaning, one can easily see how cases of divergence between (i) what the speaker (secondarily) means and (ii) what the speaker implies can be generated. A speaker irrationally believes that her audience can infer that she means that *p* in a context where her audience would infer that she means something else.

We have not yet considered the possibility that Bettina does not know that Albert knows nothing of her beliefs about aliens and aspirins. Bettina might even (falsely) believe that Albert knows. The case would then look quite different. Bettina is not irrational. She is simply mistaken about what is mutually known between Albert and herself. She thinks that her beliefs about aliens and aspirins are common knowledge between them, but they aren't. Now, I do think that Saul intends her case to involve some sort of irrationality on Bettina's part. This is at least how the case reads, given Bettina's independently outlandish beliefs about aliens and aspirins. Bettina is just the kind of person whose rationality is disputable. It is then fair to ask whether the intuitive judgement that Bettina implies something other than what she means depends on reading the case as involving irrationality on Bettina's part. If Bettina is merely mistaken about what is common knowledge between Albert and herself, what is our verdict on the case? My own judgement is still that Bettina implies *that ingesting an aspirin will soothe Albert's headache*. And so I am tempted to conclude that there are two kinds of cases of divergence between (i) what the speaker (secondarily) means and (ii) what the speaker implies: cases of speaker-irrationality, and cases of speaker mistake about what is mutually known (believed, accepted).

My judgement on one case is certainly not enough to feel confident about this conclusion. In the absence of experimental studies testing the judgements of multiple ordinary speakers, the only thing to do here is to present other cases with the same structure to test the robustness of individual judgements across cases. We need more cases in which the speaker expects her audience to use a certain proposition in the inference to what she secondarily means, mistakenly taking this proposition to be common knowledge between her and her audience.

Fortunately, such cases are not hard to build. Alf tells his friend Barb: '*Barry Lyndon* is showing at the cinema tonight. Wanna join?'. Barb answers: 'My sister's in town!'. Here is the twist: Barb's sister is a huge Stanley Kubrick fan, and Barb *falsely* believes that Alf knows this. Barb therefore means *that she will join with her sister*, and she expects Alf to recognise this, although Alf is of course in no position to recognise this. What does Barb imply? My verdict is that Barb implies *that she will not join*, and thus implies something other than what she means. To flesh out this intuition a little: given Alf's ignorance of Barb's sister's appreciation of Stanley Kubrick, it is reasonable for Alf to infer that Barb means *that she will not join.* And what is reasonable for Alf to infer is ultimately what counts when judging what Barb implies.

Here is yet another case. Art and Beth work in a law firm. Art comes to suspect that two partners of the firm plot to establish a rival firm and take some clients with them. At the office Christmas party, the mood is light and flirty. Art's memory is affected by the free prosecco, and he misremembers telling Beth about his suspicion of a corporate plot. In reality he hasn't told Beth, and Beth does not suspect anything of this sort. As the two supposedly plotting partners - call them Cedric and David - leave the office party together, Art tells Beth intently: 'So these two are leaving the party together…'. Art means something like: *they are off plotting their corporate coup*. But my verdict on this case is that Art implies something else. Art implies something like: *Cedric and David are an item*. Again, my intuitive verdict is guided by what the audience (Beth) could reasonably take the speaker (Art) to mean given what is actually mutually known between them.

It is interesting to note, as Saul does, that some of Grice's pronouncements on conversational implicatures go in the same direction:

> … what is implicated is <u>what it is required that one assume a speaker to think</u> in order to preserve the assumption that he is following the Cooperative Principle… (Grice 1989, p.86, emphasis mine)

> … an implicatum [...] is <u>the content of that psychological state or attitude which needs to be attributed to a speaker</u>… (Grice 1989, p.370, emphasis mine)

> … what I have been calling conversational implicature is just <u>those assumptions which have to be attributed to a speaker to justify him</u>… (Grice 1989, p.370, emphasis mine)

What a speaker conversationally implicates is then what her audience 'needs / is required / has' to assume the speaker means to maintain the assumption that the speaker is being cooperative. It is not obvious how 'needs to / is required to / has to' should be interpreted, but one natural interpretation is that the normativity suggested by these phrases is that of rationality. Given what the audience knows about the speaker and the context of utterance, it is reasonable for the audience to believe that the speaker means a certain proposition *p*. In general one 'needs / is required / has' to believe what it is reasonable to believe. And so the hearer 'needs / is required / has' to believe that the speaker means that *p*. Given that Grice's theoretical notion of conversationally implicating that *p* is put forward as capturing the pre-theoretical notion of implying (suggesting, hinting) that *p*, it seems that Grice would agree that Bettina implies that ingesting an aspirin will soothe Albert's headache, that Barb implies that she will not join, and that Art implies that Cedric and David are an item.

So far I have contended that there are two kinds of cases of divergence between meaning that *p* and implying that *p*. When a speaker irrationally expects her audience to make an inference the audience is in no position to make, and the context of utterance is such that another inference is reasonable, the speaker implies something other than what she means. Alternatively, when the speaker falsely believes that a proposition she expects her audience to use in her inference is common knowledge, and the context of utterance is such that another inference is reasonable, the speaker implies something other than what she means.

Implying that *p* can also diverge from *being taken by one's audience to (secondarily) mean that p*. An audience can be irrational or mistaken just as a speaker can be irrational or mistaken. Our current stock of cases can be flipped so that the audience, rather than the speaker, is at fault. A variant of *Albert and Bettina* may be built in which *Albert* has outlandish beliefs about aliens and aspirins. Albert may falsely believe that Bettina knows about his outlandish beliefs. So when he tells Bettina 'I've got a headache' and she answers 'There's aspirin in the bathroom cabinet', he infers that Bettina means that *he should flee*. But intuitively this is not what Bettina implies. Bettina means and implies that *ingesting an aspirin will soothe his headache*. Now, take a variant of *Alf and Barb* in which Barb simply means that *she will not be able to join* when uttering 'My sister's in town'. Let us further imagine that Alf falsely believes that Barb has told him about her sister's chaotic personality. In fact, Barb hasn't told him anything like this, and Barb's sister is a very collected, organised person. Because of his false beliefs, Alf infers that Barb means something like: I do not know yet whether I will join because of the usual chaos that accompanies my sister's presence. Intuitively, this is not what Barb implies. Barb implies just what she means: that she will not join.

In order to conclude that *implying that p* is public, we must further provide cases of double divergence, as we did for *saying that p*. The double divergence here is that of implying that *p* from both meaning that *p* and being taken by one's audience to mean

that *p*. Such cases may be assembled by adding audience-mistake to speaker-mistake. We can for example build a variant of *Alf and Barb* in which the two following things happen. Barb falsely believes that her sister's appreciation of Stanley Kubrick is common knowledge between Alf and herself; and Alf falsely believes that the chaotic personality of Barb's sister is common knowledge between Barb and himself. Here is the dialogue again:

Alf: *Barry Lyndon* is showing at the cinema tonight. Wanna join?
Barb: My sister's in town!

Barb means that she will join with her sister. Alf takes Barb to mean that she does not know yet if she can join, due to her sister's chaotic personality. I still think that in this case what is implied is that *Barb will not be able to join*. After all, this is what a reasonable audience who is not mistaken about the mutual knowledge between Alf and Barb would take Barb to mean. I also admit that my intuitions become more muted in such cases, compared to cases of single divergence. This weakness in intuitions mirror those on *what is said* in cases of double factual mistake (speaker-mistake plus audience-mistake), as explained in section 1.a. When the speaker errs but the audience puts up an impeccable interpretive performance, we side with the audience. When the audience errs but the speaker performs an impeccable communicative performance, we side with the speaker. But when both speaker and audience err, we have no one left to side with. If we really have to decide on a propositional content for *what is implied*, we have to think about what an audience that makes no mistake about what is mutually known would infer. This might come close to Grice's own characterisation of what is *implicated*, as already noted: *"... what is implicated is what it is required that one assume a speaker to think..."* (Grice 1989, p.86).

Despite this reservation, it seems to me that there are cases in which what a speaker implies diverges from both what she means and what her audience takes her to mean. If cases of double divergence of *what is implied* do exist, one may conclude that what a speaker implies is *public*. Now, how does the publicity of what is implied bear on the semanticity of what is said? The data we have to explain is the following: both the content of say-judgements and the content of imply-judgements are public. There are possible explanations of this data that involve the semanticity of what is said. Indeed, one can provide a *disunified* explanation: the content of say-judgements is public because it is semantic, while the content of imply-judgements is public for some other reason. Other possible explanations involve the *over-determination* of the publicity of what is said. According to an over-determination view, the content of say-judgements is public because it is semantic, but even if it was not semantic it would still be public in virtue of something else. This something else independently explains why what is *implied* is public.

A unified explanation of the overall data demands that one reject a semantic explanation of the publicity of what is said. A unified explanation tells us that both the content of say-judgements and the content of imply-judgements are public *in virtue of the same facts*.

Since the publicity of the content of imply-judgements cannot be explained by its semanticity, the publicity of the content of say judgements will not be explained by its semanticity either.

Other things being equal, a unified explanation is superior to a disunified explanation. A unified explanation is also superior to an explanation that posits over-determination. My conclusion is therefore conditional: if there is a plausible unified explanation of the publicity of what is said and what is implied, one should favour this explanation. I should further note that disunified explanations and over-determination explanations must account for the publicity of what is implied anyway, and that this explanation cannot be that what speakers imply is semantic. Hence, proponents of disunified explanations and over-determination explanations are saddled with an additional explanatory burden: they must explain why the non-linguistic explanation of the publicity of what is implied is not good enough for the publicity of what is said.

I do not attempt to provide a unified explanation of publicity in the remainder of this chapter. For this, one will have to wait until the final chapter of this dissertation. In the next section, I present other properties of the content of say-judgements which allegedly support its semanticity, before arguing that they do not support it.

2. Productivity* and semanticity

a. Productivity and productivity*

Jeff Speaks has recently argued that certain properties of our knowledge of the reference of demonstratives in context are best explained by the idea that demonstratives have an object-determining rule as their meaning - something like a Kaplanian character (Speaks 2017, pp.712-714). For the purposes of this chapter, Speaks' claim may be transposed as the claim that certain properties of pre-theoretical say-judgements on utterances of sentences containing demonstratives are best explained by the semanticity of what is said.

The properties of say-judgements that Speaks invokes are analogues of *productivity* and *systematicity*. Here I focus on productivity and its analogue. Productivity is a property of our knowledge of natural languages such as English or French. More specifically, productivity is a property of our knowledge of the meaning of expressions of these languages. Someone who speaks a language is able to understand a great number of sentences - or complex expressions more generally - she has never encountered before. This property is commonly thought to support the compositionality of meaning in natural languages, i.e. the thesis that the meaning of complex expressions is determined by their syntactic structure and the meaning of their constituents (Pagin & Westerstahl 2010a, 2010b; Szabo 2020). The line of thought goes like this. If I speak English, then I understand 'Cats swallow peas' although I have never encountered this sentence before. How is this possible? Of course, part of what makes it possible for me to understand 'Cats swallow peas' is that I know what 'cats', 'swallow', and 'peas' mean. Knowledge of the syntactic structure of 'Cats swallow peas' - [$_{S}$ [$_{NP}$Cats] [$_{VP}$

[$_{V}$swallow] [$_{NP}$peas]]] - is also necessary. But these two things are not sufficient: I must arrive at the meaning of the sentence from the meaning of the words and the syntactic structure of the sentence which contains them. According to friends of compositionality, I can do this as a speaker of English because my knowledge of English contains rules of assembly for the meanings of simple expressions given the syntactic structure of the complex expressions in which the simple expressions occur. In other words, my knowledge of English includes *rules of meaning-composition.*

One can then hypothesise that knowledge of a language consists in knowledge of the meaning of its simple expressions, knowledge of the syntactic structures in which they can occur, and knowledge of rules of meaning-composition. A theory of linguistic knowledge should then put forward syntactic structures and meanings which are consistent with the fact that meanings compose according to rules. This theory should also postulate specific rules of meaning-composition, and demonstrate how these rules yield complex meanings that match pre-theoretical judgements on sentences.

Now, Speaks introduces the further property of productivity* (notation mine). Productivity* is supposed to be a property of our knowledge of the content of context-sensitive expressions. The knowledge in question is productive* in the following sense: whenever a competent speaker of a language is given a novel pair consisting of a sentence containing a context-sensitive expression plus a context of utterance, she is able to judge what is said with this sentence in this context. In other words, competent speakers of a language know what is said when they encounter a novel sentence-context pair.

Speaks builds an analogy between productivity and productivity*.[59] The analogy is supposed to work as follows:

Productivity:

- Knowledge of the syntactic structure of a sentence + knowledge of the meaning of the constituents of the sentence -> knowledge of the meaning of the sentence.

Productivity*:

- Knowledge of the context of utterance + knowledge of the meaning of the sentence -> knowledge of what is said.

From this analogy, Speaks takes a leap. As I have explained, productivity is usually invoked in an argument to the best explanation, the best explanation being that our linguistic knowledge includes rules of meaning-composition. This argument is widely accepted in

[59] Speaks further assumes that knowledge of the *content* of context-*insensitive* expressions is productive (no star), which supports a compositional theory of content for the context-insensitive part of our language (Speaks 2017, pp.711-714). Speaks' analogy is then strictly speaking one between the productivity* of knowledge of content on the one hand, and the productivity of knowledge of content (for context-insensitive expressions) on the other. I ignore this aspect of Speaks' argument for several reasons, the main reason being that Speaks' argument does not stand or fall on the productivity of knowledge of *content* (as opposed to knowledge of *meaning*). As I see it, the argument fails either way.

linguistics and philosophy. According to Speaks, productivity* may then be invoked in an analogous argument to the best explanation, the best explanation of productivity* being that our linguistic knowledge includes rules associated with context-sensitive expressions that determine their content *via* certain features of the context of utterance. The two analogous *explanantia* are thus meaning-composition rules (in the case of productivity), and content-determining rules for context-sensitive expressions (in the case of productivity*). Productivity* is then supposed to show that all context-sensitive expressions, including demonstratives, have something like a Kaplanian character. If this is right, then what is said with sentences containing context-sensitive expressions is *semantic*. Speaks' argument from productivity* is then an argument for the semanticity of the content of say-judgements in general, including say-judgements on sentences containing context-sensitive expressions.

According to Speaks, the structural similarity between the two arguments - i.e. the argument from productivity and the argument from productivity* - and the widespread endorsement of the former argument in the philosophical community are enough to endorse the latter argument. As he puts it, *"if this reasoning is persuasive in that case, it is hard to see why it should not be so in the present case as well"* (Speaks 2017, p.713). Hence, Speaks does not argue directly that the semanticity of what is said is the best explanation of productivity*. He relies on the argument from productivity and the structural similarity between the two arguments.

In their (2005), King and Stanley make a related case for the semanticity of the content of say-judgements. King and Stanley take our knowledge of what is said with sentences ('what is expressed') to possess a property I shall call *productivity***.

Productivity**:

- Knowledge of the syntactic structure of a sentence + knowledge of the meaning of the constituents of the sentence + knowledge of the context of utterance -> knowledge of what is said.

Unlike Speaks, King and Stanley do not bother to go through an analogy with productivity. They argue directly from the productivity** of what is said to its semanticity. In their own words:

> Given a finite vocabulary, and grasp of the composition rules expressed by syntactic structures, speakers have the ability to grasp the propositions expressed by an infinite number of sentences. If language users employ a compositional semantic theory in grasping the contents of speech acts, then one has a satisfactory explanation of the systematicity and productivity of a speaker's grasp of an infinite number of novel utterances. For then one can explain a language user's grasp of what is expressed by the utterance of a novel sentence by appealing to the fact that she grasps the words in the sentence and their modes of combination, <u>together with whatever contextual information is required to</u>

interpret the context-sensitive elements in the sentence. (King and Stanley 2005, p.140, emphasis mine)

In the next section I show that these arguments for semanticity from productivity* and productivity** fail.

b. From productivity* to semanticity?

Speaks argues that the structural similarity between the argument from productivity and the argument from productivity*, combined with the widespread acceptance of the former argument, justifies endorsing the latter argument. He seems confident about the overall shape of this meta-argument: "*it is hard to see how [*the conclusion of the argument from productivity*] could be reasonably rejected, given that our knowledge of the contents of indexicals in arbitrary contexts seems to be parallel in the ways detailed above to our knowledge of complex expressions"* (Speaks 2017, p.732). Granting for the sake of argument that there is nothing wrong with the shape of Speaks' meta-argument, this meta-argument should however be weighed against the independent plausibility of an abductive argument from productivity* to semanticity.

Taken on its own, the argument from productivity* to semanticity is an argument to the best explanation. Of course, there are alternatives to an explanation of productivity* in terms of content-determining linguistic rules. Here is one alternative explanation. It is our pragmatic competence - our ability to recognise the communicative intentions of speakers - that allows us to know what a speaker says once we know the meaning of the uttered sentence and the context of utterance. It is uncontroversial that we have a competence of this sort on top of our linguistic competence, given how much interpreters are able to retrieve from an utterance (including what the speaker implies) and how much speakers intend their interpreters to retrieve. Now, suppose that no content-determining linguistic rule is associated with context-sensitive expressions such as demonstratives. Given our pragmatic competence, we are still able to associate a proposition with a novel sentence-context pair. We are able to use the sentence's meaning and the context of utterance as premises in an inference to the communicative intentions of the speaker.[60] It is then no wonder that we get knowledge of what someone says from knowledge of the meaning of the uttered sentence and knowledge of the context of utterance. This picture according to which knowledge of what people say is based partly on linguistic competence and partly on pragmatic competence has its contemporary proponents (Bach, 2017, Harris 2020, Heck 2014). This pragmatic explanation of productivity* is also an explanation of productivity**.

[60] I am not suggesting that the existence of a content-determining rule is inconsistent with the intervention of our pragmatic competence in content-ascription. After all, some expressions could have a character that appeals to the communicative intentions of the speaker. The point is only that without a character our pragmatic competence would spring into action anyway to associate a content with the expression (see Chapter 2 section 4).

Of course, given my further view that the content of pre-theoretical say-judgements is *public*, the alternative pragmatic explanation is not a straightforward option for me. If what a speaker says is public, it is not a matter of what she means. And so it seems that pragmatic competence, which serves to ascertain what speakers mean, should not be involved in ascertaining what speakers say. This appearance may be resisted though. Without going into too much detail, a plausible explanation of productivity*/productivity** could be given that makes a qualified appeal to pragmatic competence. Consistently with publicity, if what people say is roughly what a non-actual audience with certain properties would be justified to take the speaker to mean, figuring out what the speaker said could involve some kind of *pragmatic simulation*. The idea is that we first 'put ourselves in the non-actual audience's shoes' and then deploy our pragmatic competence from the perspective of this audience.

Now, the simple fact that there is an alternative explanations of productivity* does not on its own undermine an abductive argument from productivity* to semanticity. As long as semanticity is a better explanation than the alternatives, the argument goes through. However, the fact that no analogous pragmatic explanation of *productivity* can be given should give one pause. Suppose that our linguistic knowledge does not include rules of meaning-composition. Our linguistic knowledge only consists of knowledge of grammar and knowledge of the meaning of simple expressions. A *pragmatic* explanation of productivity would have it that our pragmatic competence comes into play in the assembly of simple meanings into complex meanings given a certain syntactic structure. This explanation would be wildly implausible. First, interpreting utterances of complex expressions would be cognitively costly: every time we would encounter a complex expression, we would need to come up with a composition mechanism that leads to a complex meaning that matches the speaker's communicative intentions. Composition would not be 'automatic'. Secondly, something like a category mistake is involved in the idea that our pragmatic competence can yield rules of composition. The domain of our pragmatic competence is certain mental states of the speaker: her communicative intentions, and further mental states embedded in these intentions (e.g. the belief that *p*). Rules of composition are not part of this domain.

The fact that there is no plausible pragmatic explanation of *productivity* makes an abductive argument for rules of meaning-composition much more straightforward than one for object-determining rules for context-sensitive expressions. That our linguistic knowledge includes rules of meaning-composition is *the only* plausible explanation of productivity. By contrast, that our linguistic knowledge includes object-determining rules is one plausible explanation among several plausible explanations of productivity*/productivity**. An abductive argument for the semanticity of what is said from productivity*/productivity** must then include reasons to prefer a semantic explanation to a pragmatic one. *Contra* Speaks, the analogy between the argument from productivity and the argument from productivity* plus the widespread acceptance of the former are not enough. One must show that the semantic explanation of productivity* really is the best one, since it is not the only plausible one.

Crucially, there is a reason to think that the semantic explanation of productivity*/productivity** is not the best one. Productivity* and productivity** also hold of our knowledge of contents which are uncontroversially not semantic. For example, productivity* and productivity** also hold of our knowledge of *what speakers imply*:

Productivity*:

- Knowledge of the context of utterance + knowledge of the meaning of uttered expressions -> knowledge of what the speaker *implies.*

Productivity**:

- Knowledge of the syntactic structure of a sentence + knowledge of the meaning of the constituents of the sentence + knowledge of the context of utterance -> knowledge of what the speaker *implies.*

When we encounter a novel utterance of a sentence, knowing the context of utterance and knowing the meaning of the uttered sentence is enough to know what the speaker implies. When I presented *Alf and Barb* to the reader - a novel utterance - the reader knew what Barb implied. The reader knew this on the basis of the meaning of the uttered sentence and the context of utterance. Whereas an *analogy* holds between our knowledge of the meaning of complex expressions and our knowledge of what people say, a true *parallel* holds between our knowledge of what people say and our knowledge of what people imply:

Productivity*:

- Knowledge of the context of utterance + knowledge of the meaning of expressions -> knowledge of what is *said*.
- Knowledge of the context of utterance + knowledge of the meaning of expressions -> knowledge of what is *implied*.

Productivity**:

- Knowledge of the syntactic structure of a sentence + knowledge of the meaning of the constituents of the sentence + knowledge of the context of utterance -> knowledge of what is *said*.
- Knowledge of the syntactic structure of a sentence + knowledge of the meaning of the constituents of the sentence + knowledge of the context of utterance -> knowledge of what is *implied*.

Since the productivity*/productivity** of our knowledge of what is implied cannot be explained in terms of content-determining linguistic rules, a unified explanation of the productivity*/productivity** of both knowledge of what is said and knowledge of what is implied will not appeal to content-determining linguistic rules.

A predictable reply to my line of reasoning about productivity*/productivity** is that the parallel I establish between knowledge of what is said and knowledge of what is implied equivocates on 'context of utterance'. According to this reply, different facts are targeted by 'context of utterance' in the case of what is said and in the case of what is implied. In the case

of what is said, only a short list of facts is relevant: who the speaker is, what the time is, where the speaker is, and perhaps a few extra features. In the case of what is implied, *any extra-linguistic fact* may be relevant. In *Alf and Barb*, it is clear that what is mutually known between Alf and Barb is relevant to what Barb implies. By contrast, what is mutually known between Alf and Barb is not relevant to what Barb *says*. Or so the reply goes.

My counter-reply is simple: there is no restrictive notion of a context of utterance that is exclusively relevant to what speakers say. Contrary to what the reply would have us believe, any extra-linguistic fact might be relevant to determining what a speaker says, just as any extra-linguistic fact might be relevant to determining what a speaker implies. Heck has made a convincing case for this in a recent article, with some delightful examples (Heck 2014). Here are two of them:

> Some time ago, I was at the aquarium with my daughter when she said, "That's a pretty fish". I looked up and there, high in the corner, was a beautiful blue and yellow fish. I then knew to which fish she had referred, and I knew it every bit as surely as if she had pointed directly at that fish. It wasn't that it was the most beautiful fish in the tank—though it might have been, and then that would be a different example. In this case, rather, what led me to the right fish was the fact that I knew that this was a fish that Isobel, in particular, would think was pretty. (Heck 2014, pp.338-339)

> Some years ago, I was driving with my friend "Steve" following a pleasant lunch out. About twenty minutes after we left the restaurant, Steve said, just "out of the blue," "She was gorgeous." I knew immediately to whom he was referring: the hostess at the restaurant where we'd just had lunch, whom I shall call Sarah, and who looked as if she'd just arrived from a *Vogue* photo shoot. [...] As I said, I knew immediately to whom Steve was referring, but that is because I had been at lunch with him, and I too had noticed how strikingly beautiful Sarah was—something he expected me to have noticed. But of course, it is not at all essential that she actually was gorgeous. I would have known just as well to whom he was referring had it just been clear that he was smitten by her. Or she might have been the sort of woman to whom I knew he was particularly likely to be attracted. (Heck 2014, pp.337-338)

In the first example, Isobel says of the blue and yellow fish high in the corner that it is a pretty fish. It is not possible to know what Isobel says without knowing *what kind of fish Isobel finds pretty*. This fact certainly does not belong to the short list of extra-linguistic facts that the aforementioned reply would have us believe is exclusively relevant to what is said. As for the second example, Steven says that Sarah was gorgeous, but it is not possible to know what Steven says without knowing that *Sarah is gorgeous*, or without knowing *what kind of woman Steven finds gorgeous*. Again, these facts do not belong to the short list of extra-linguistic facts which some theorists think are exclusively relevant to what is said.

Since a true parallel can be drawn between the productivity*/productivity** of our knowledge of what is *said* and the productivity*/productivity** of our knowledge of what is *implied*, the theorist is under pressure to provide a unified explanation of the overall phenomenon. And this unified explanation cannot appeal to the semanticity of what is said.

Conclusion

The content of pre-theoretical say-judgements sometimes diverges from both what the speaker intends and what her audience takes her to intend (section 1.a.). The possibility of double divergence if this sort supports the *publicity* of what is said: what is said is neither determined by the speaker's intentions nor by the audience's interpretation (section 1.b.). The publicity of what is said is easily explained by its semanticity, but the fact that uncontroversially non-semantic contents such as what is implied are also public undermines this explanation (section 1.c.). Our knowledge of what is said productive*/productive**. These further properties are explainable by the semanticity of what is said (section 2.a.). But our knowledge of uncontroversially non-semantic contents such as what is implied also displays these properties, which undermines the semantic explanation (section 2.b.).

Chapter 6: A modal analysis of defective content-acts

In the previous chapter, I argued that if there is a plausible unified explanation of the publicity of both what is said and what is implied, then one should favour such an explanation. I will provide this explanation in the next chapter. In order to do so, I must first make an intermediate move. In this chapter, I analyse an intuitive defect that attaches to utterances in cases of unintentional divergence between what the speaker says and what she means. I propose that this defect consists in the fact that the audience would not recognise what the speaker means. Following King's lead, I further elucidate this modal condition by appealing to Stalnaker's notion of common ground (King 2013, 2014; Stalnaker 2002, 2014). This elucidation, and modal conditions on the interpretation of a non-actual audience more generally, have been recently criticised (Nowak & Michaelson 2021, Peet 2021, Speaks 2016). I show that my preferred elucidation of the modal condition is more resilient than the literature would have us believe.

1. An intuitive defect in utterances

In the Carnap-Agnew case, there is something wrong with the speaker's communicative act. The speaker intends to communicate a proposition about Carnap's picture to her audience by pointing at it, but she ends up pointing at a different picture. Intuitively, the speaker puts up a poor communicative performance. It is important to note that the defective character of the speaker's performance does not consist in the failure of communication. In fact, it seems to be independent from it. The original Carnap-Agnew case is under-described by Kaplan: one does not know who the audience is, what the audience knows, etc. Hence, one does not know from Kaplan's presentation of the case if communication fails. Yet one gets from this incomplete presentation an intuition that something is wrong with the speaker's communicative act. Hence, the intuition of a defective communicative performance does not seem to depend on the success or failure of communication.

Now, it is possible that we get this intuition in a case as patchy as the Carnap-Agnew case because we fill out the details of the case in a specific way. Maybe we conjure up a listener who has no idea about Carnap, or about the picture of Carnap usually hanging on the wall. Of course, *this* listener takes the speaker to be talking about the picture behind her - Agnew's picture - so communication fails. But here is the crucial twist. Even when we fill out the details of the case in a different way, so that the listener recognises that the speaker is talking about Carnap's picture (because she knows about Carnap's picture and the speaker's obsession with Carnap, or because she knows about the picture switch, or for some other reason), there is still a lingering intuition that the speaker's communicative performance is defective. If we trust this intuition, we must acknowledge that a *successful* communicative act can be defective. The intuited defect seems to be independent from the actual success or failure of communication.

The Carnap-Agnew case is a case of unintentional divergence between what the speaker primarily means and what she says. Many other cases of unintentional divergence seem to elicit a similar intuition of deficiency. I will not go through every kind of case of unintentional divergence introduced in the previous chapter, but let me just note for the sake of illustration that cases of divergence due to a *linguistic mistake* or a *linguistic accident* elicit the relevant intuition. When a speaker utters 'Bob is evasive' intending to communicate that Bob is elusive (either because she's mistaken about the conventional meaning of 'evasive' or because her tongue slips), we not only get the intuition that she said that Bob is evasive, but also that she put up a poor performance. I allow myself to hypothesise that the unintentional divergence between what is meant and what is said is somehow related to the intuitive deficiency of the speaker's performance, and that all cases of such unintentional divergence elicit the relevant intuition.

I further assume that the defect in cases of divergence involving a mistake is the same as the defect in cases of divergence involving an accident. This does not quite follow from previous assumptions. But if on the one hand both mistakes and accidents cause unintentional divergence, and on the other hand all cases of unintentional divergence elicit an intuition of defective speaker-performance, it is natural to suppose that the intuited defect is the same in cases of accident and in cases of mistake. This assumption rules out an account of the defect in terms of mistake (or in terms of accident). It forces us to hypothesise that a feature common to cases involving a mistake and cases involving an accident is responsible for the defect intuitively detected in both kinds of cases.

I am not the first to discuss an intuited defect in the speaker's performance in cases of unintentional divergence. It has notably been discussed in the metasemantics for demonstratives literature. Cases of unintentional divergence between intended reference and actual reference such as the Carnap-Agnew case have been put forward as counterexamples to intentionalism. Such cases elicit an intuition of defective communicative performance on the part of the speaker. Here are some characterisations of the intuited defect by metasemanticists:

> *...in these cases, the speaker has not done enough to put her audience in a position to discern her referential intention.* (Speaks 2016, p.309)
>
> *Intuitively, the speaker failed to discharge her responsibility to be understandable.* (King 2014, p.225)
>
> *If the speaker fails to make his* [intended] *reference available, his speech act is defective, and not even the best intentions can repair the defect.* (Wettstein 1984, p.75)

As I see it, these various formulations have two things in common. First, they all suggest that the intuited defect does not consist in actual communicative failure. Secondly, they all suggest

that the intuited defect is yet somehow closely related to the possibility of communicative success. The speaker does not 'put her audience in a position to discern her referential intention'; she is not 'understandable'; and she fails to 'make her intended reference available'. At first glance, these formulations converge towards the idea that the speaker's audience cannot retrieve the intended referent, and hence that communication cannot succeed. But 'cannot' cannot be right. The impossibility of communicative success entails actual communicative failure, and I have already argued that the intuited defect can be present even in cases of actual communicative success. As a matter of fact, the cited authors do not take the intuited defect to consist in the *impossibility* of communicative success. Showing some flexibility about the modality in the above formulations, one might say that the intuited defect consists in the fact that *the audience would not understand what or who the speaker intends*. In other words, communication fails in some range of non-actual possibilities. This is in fact how Jeff King wishes to develop the idea that the speaker is 'not understandable' in such cases (King 2013, 2014). Following King, I will attempt to analyse the intuited defect with a modal condition on the interpretation of the speaker's utterance.[61]

Before turning to the laborious task of fleshing out a modal condition, let me note that a condition of this sort validates the assumption that the intuited defect is the same in cases of mistake and in cases of accident. In both types of cases, the speaker would be taken to mean something other than what she means. As I see it, a condition on what the speaker would be taken to mean is one way to capture the idea that the intuited defect has to do with the speaker's utterance *not furthering communicative success*. Speakers devise an utterance that they think will allow their communicative goals to be fulfilled. They devise a plan involving certain communicative means to attain a certain communicative end. But sometimes the communicative means speakers end up employing do not in fact further their communicative end. In this regard, both an ill-conceived plan to attain an end and a poor execution of a well-conceived plan compromise the attainment of the end. In other words, both mistake and accident put communicative success in jeopardy. The 'furthering' relation between communicative means and communicative end (or the 'compromising' relation, which is the other side of the same coin) may be analysed in modal terms: given the deployment of certain means - given a certain utterance - the end would not be attained - the speaker's meaning would not be recognised.

One more qualification is needed. I have thus far spoken loosely of 'communicative act' and 'communicative performance'. I use 'communicative act' as roughly synonymous with 'speech act'. Both phrases have their drawbacks. 'Communicative act' does not rule out

[61] Both Speaks in published work (Speaks 2016, pp.330-332) and Jessica Brown in comments on a draft of this chapter have suggested that the intuitive defect can be explained by the fact that the speaker 'would not be justified in believing' (Speak's formulation) or 'cannot reasonably expect' (Brown's formulation) that her audience will recognise her intended content. This criterion sounds intuitively plausible, but a *prima facie* reason not to explore it is that it seems to make the wrong prediction in a central case for us, namely the Carnap-Agnew case. In some natural sense, the speaker of the Carnap-Agnew case can reasonably expect her audience to recognise that she means something about Carnap's picture (after all, she has no reason to believe that her picture of Carnap has been replaced).

non-linguistic acts of communication, although here I concentrate on linguistic communicative acts. 'Speech acts' superficially rules out non-linguistic actions, and it superficially rules in actions involving speech that have nothing to do with communication - e.g. clearing one's voice (Green 2013). In any case, something more precise than 'communicative act' or 'speech act' is needed for the defect to be analysed. A central tenet of speech act theory is that several acts can be abstracted from a total speech act. An illocutionary act is a matter of putting forward a certain content - e.g. the proposition *p* - with a certain force - e.g. *asserting* that *p*. From this illocutionary act one can abstract the act of putting forward a content. Austin called this the 'locutionary' act, Searle called it the 'propositional' act (Austin 1975, Searle 1968). Disagreements underlie this divergence in terminology, but I will not discuss them here. As for myself, I prefer 'content-act' to 'locutionary act' because for Austin locutionary acts are closely related to the conventional meaning of the words uttered by the speaker.[62] Since my goal is to explain the publicity of not only what is said but also what is merely *implied*, I should show foresight and choose another terminology than 'locutionary'. I also prefer 'content-act' to 'propositional act' out of foresight: I hope that what works for what is said and what is implied also works for what is promised, requested, ordered, etc. Since the latter contents might not be propositional, I should choose more neutral terminology than 'propositional'.

The defect that plagues utterances in cases of unintentional divergence between intended content and actual content affects the speaker's illocutionary act by affecting the *content-act* more specifically. Intuitively, the speaker does not fail to make clear what she is doing with the content she is putting forward. She rather fails to make clear *what content* she is putting forward. Beyond this intuitive contrast, cases of divergence can be modified so as to involve different illocutionary acts. The speaker of the Carnap-Agnew case might point behind her and say: 'Does that bother you?'. Intuitively, she has asked whether Agnew's picture bothers her audience, although she intended to ask something about Carnap's picture. Her utterance elicits the same intuition of deficiency as in the assertoric case. The speaker might alternatively point behind her and utter 'Get that out of the room!'. We still get divergence between intended content and actual content, plus an intuition of defective performance. Deficiency remains as illocutionary force varies. Hence, the intuited deficiency pertains to the content-act within the illocutionary act. The illocutionary component of speech acts and its potential defects will not be discussed in this chapter. I will use 'performance' and 'act' when convenient, but one should remember that it is the content-act within the illocutionary act that is at stake.

2. In search of the right modal analysis of the defect

[62] Another reason not to adopt the 'locutionary act' terminology is because of a relatively recent paper by Braun, in which the author defines *locuting a content c* as the act of uttering a linguistic expression that semantically expresses *c* (Braun 2011). A locutionary act is then the act of putting forward a content, but this content is tied to the conventional meaning of the expression uttered by the speaker. My notion of a *content-act* is not tied to conventional meaning in this way.

We start with a rough modal condition: the content-act performed by a speaker is defective just when the audience would not retrieve the speaker's intended content. As with any modal condition M*p* - where M is a modal operator and *p* is a proposition - the difficulty is to give the meaning of the modal operator M. And to give the meaning of M for a condition M*p* is to specify the range of possibilities in which *p*. In the present case, we must specify the range of possibilities in which the audience does not retrieve the speaker's intended content. We must do so in such a way that the condition yields correct predictions: predictions of deficiency when the speaker's act is intuitively deficient, and predictions of non-deficiency when the speaker's act is intuitively faultless.

Some choices here are obvious, and some less obvious. Intuitively, the idea with our modal condition is that the audience would not recognise what the speaker means given (i) the utterance produced by the speaker; (ii) the communicative intentions underlying this utterance; and (iii) the context of utterance. So we know from the start that the words and gestures actually produced by the speaker, as well as her communicative intentions, partly define the relevant range of possibilities - the *modal domain*. Additionally, if we think about Carnap-Agnew-style cases, it seems that the actual context of the utterance must be held fixed across the modal domain. What makes the speaker's act defective in the Carnap-Agnew case is that Agnew's picture is behind the speaker when she produces her utterance. Remove this feature of the case and there is no case to speak of.

Beyond these easy stipulations, it seems to me that there are three choice points with regard to the elucidation of a modal condition on the interpretation of the audience. The first choice point concerns the operative notion of context when we demand that the actual context of utterance partly define the modal domain. The second choice point concerns the strength of 'would': does 'would' express universal quantification over the modal domain, or something weaker? The third choice point concerns the nature of the audience in the modal condition. Assuming that the actual audience consists of just one person, is the audience across the modal domain the same person as in the actual world? Does it have the properties that the actual audience actually has?

Let me address each of these issues in turn. The intuition that in the Carnap-Agnew case the speaker's act is defective because Agnew's picture hangs on the wall behind her suggests that the notion of context relevant to our modal condition is *factual* rather than *mentalistic*. Context in the factual sense is the situation of the utterance or some selected features of this situation - *à la* Kaplan (1989). By contrast, context in the mentalistic sense is not made up of facts but rather of propositions accepted (or known, or believed) by the participants of a

conversation - *à la* Stalnaker (Stalnaker 1999, 2002, 2014).[63] The Carnap-Agnew case superficially suggests that context *qua* set of selected features of the situation of utterance (including for instance what the speaker demonstrates) could do for our modal condition. But this cannot be right, and this cannot be right in view of very general considerations. As we saw at the end of the previous chapter, any fact whatsoever can be relevant to recognising what a speaker primarily means. If the speaker counts on her audience drawing on a fact outside the small circle of Kaplanian selected features, then this fact should obtain across the modal domain expressed by our modal operator. Now, if any fact can be relevant to retrieving the speaker's intended content, does this require that we hold fixed across the modal domain *every* fact that actually obtains? This cannot be right either. A modal domain made up of possibilities indistinguishable from actuality is in effect a domain containing the actual world only. Our modal condition would thus reduce to a condition of actual communicative failure, which is exactly what we do not want.

The context of utterance suited to our purposes is such that: (i) any fact can be part of it, but (ii) it is not identical with the totality of facts. It seems to me that Stalnaker's view of the context of utterance as the set of mutually accepted propositions between participants of a conversation - the 'common ground' - comes closest to satisfying these desiderata. Switching from *facts* to *propositions*, one can make sense of the idea that any *proposition* can be relevant to discerning the speaker's intended content (since any proposition can be part of the common ground of the conversation), without marching towards the conclusion that the context relevant to our modal condition is the whole world of the utterance. However, adopting a Stalnakerian notion of context would mean revising the initial thought that the context suited to our purposes is *factual*. A context of utterance *qua* common ground is definitely *mentalistic*. In what follows I give common ground a chance: we shall judge this conception of the context of utterance by its fruit.

Stalnaker claims that the context supposed to determine the content of a speaker's utterance is not the common ground *prior to the speaker's utterance* (Stalnaker 1999, p.86). Otherwise, no utterance of sentences containing 'I' or 'You' could ever yield successful speech acts. My utterance of 'I' presupposes that I am the speaker, but the proposition that I am the speaker cannot be common ground before I open my mouth. Is the same true of the context that partly defines the range of possibilities expressed by the 'would' of our modal condition? Here is one reason to think not. Suppose that unbeknownst to the speaker her audience becomes inattentive as she is speaking. That the speaker uttered the sentence she uttered, and that she made the gestures she made does not make it into the post-utterance common ground because

[63] It might seem misleading to speak of a Kaplanian context as a small set of selected features, since Kaplan's contexts include the *world* of the utterance. However, as I understand Kaplan's work, the motivation for including the world of the utterance as part of the context of the utterance is not that what is said can be determined by any fact whatsoever, and thus that the context of the utterance includes the world of the utterance *qua* totality of facts. When it comes to the determination of what is said, the world parameter in Kaplan's contexts is there only to fix the content of 'actually'. The world parameter plays a further theoretical role for Kaplan that has nothing to do with the determination of what is said. It serves to make contexts *proper*, which in turns allows to define validity in his logical system in terms of truth in every context.

of the audience's inattention. And no audience would retrieve the speaker's intended content from *this* common ground. Hence, if the actual *post-utterance* common ground defines our modal domain, our modal condition predicts that the speaker's performance is defective. Intuitively, this is the wrong result. A speaker's performance is not defective when she speaks while, unbeknownst to her, her audience has lost her focus and stopped listening. The context that obtains across the range of possibilities targeted by the 'would' in our modal condition is thus the common ground of the conversation *prior to the speaker's utterance*.

The second issue raised by a rough modal condition is whether the condition *that the audience does not retrieve the speaker's intended content* must be fulfilled in every possibility in the modal domain, or only in some subset of these possibilities. Does 'would' express universal quantification over the modal domain, or does it express something weaker? Here are two reasons to go for the latter option. First, take any case in which the speaker's utterance makes it hard for her audience to discern who she is referring to - e.g. a vague pointing gesture towards a crowd of men accompanied by an utterance of 'He'. We want to say that the speaker's content-act is defective, and thus that the audience would not retrieve the speaker's intended referent. But the audience might venture a guess about the intended referent to break the interpretive impasse. Unless an explicit provision bans such guesses from the modal domain, there will be possibilities in the modal domain in which the audience ventures a correct guess about the speaker's intended referent. If our modal condition for a defective content-act requires that the audience does not retrieve the speaker's intended content in any possibility in the domain, then our modal condition will incorrectly classify such cases as cases of faultless speaker-performances. Secondly, consider a symmetric modal condition for *non-defective* content-acts: a content-act is non-defective just when the speaker's audience would retrieve the speaker's intended content. An alarm could go off when the speaker is uttering her sentence, the hearer could suffer a stroke, or be struck by lightning. Unless freak events of this sort are explicitly banned from the modal domain, there will be possibilities in the modal domain in which the audience does not retrieve the speaker's intended content. If the modal condition for non-defective content-acts demands that the audience retrieve the speaker's intended content in *every* possibility in the modal domain, the condition overgenerates defective content-acts. There might be several ways to deal with these issues, but the way that seems most natural to me is to let 'would' express non-universal quantification over the relevant range of possibilities. 'Would' may be interpreted as: in *most* possibilities in the modal domain.

The third question raised by a rough modal condition concerns the audience that figures in it. Consider the simple case of an audience consisting of one person. Do we mean 'the audience' to fall outside the scope of the modal operator, in which case the interpreter across our range of possibilities is the same person that interprets the speaker's utterance in the actual world? Or do we mean 'the audience' to fall under the scope of the modal operator, in which case the identity of the audience is allowed to vary across the modal domain? The second option does not seem desirable. First, if the identity of the interpreter is allowed to vary without further restrictions on her properties, we shall find in our modal domain possibilities in which the interpreter is just terrible (linguistically incompetent, unreasonable, etc). But intuitively

again, the interpretation of a terrible non-actual interpreter is irrelevant to the quality of the speaker's performance. In addition, utterances of sentences containing 'You' are problematic for the view that the audience's identity varies across the modal domain. Assuming the identification of audience with *addressee*, and assuming that a speaker uttering 'You' means a singular proposition containing the *actual* addressee, letting the identity of the addressee vary across the modal domain seems to yield the unacceptable result that no utterance of 'You' can be faultless, since the speaker's meaning remains the same across the modal domain while the audience's interpretation varies.[64]

Now, the first option - i.e. making the audience across our modal domain identical with the actual audience - does not seem desirable either, at least if the audience is not only the same person as in the actual world but also instantiates all the properties that the actual audience actually instantiates. Suppose that, unbeknownst to the speaker, the actual interpreter is inattentive or badly mistaken about the common ground of the conversation. If this interpreter *as she actually is* is held fixed across our modal domain, then our modal condition predicts that the speaker's act is defective. Again, the speaker is intuitively not responsible for unforeseeable shortcomings of her audience, and these shortcomings do not automatically make the speaker's act defective.

We can find a way out of this impasse in the metasemantics for demonstratives literature. Several authors allow the possible interpreter to have properties that the actual interpreter does not actually have, while holding the properties of the possible interpreter fixed across the modal domain (Wettstein 1984, King 2014, Speaks 2016). I cannot discern whether these authors take the possible interpreter to be *the same person as* the actual interpreter. This does not seem to matter to them: it is the properties of the interpreter other than her identity that matter. Given the problem raised by 'You', I would suggest that the interpreter across our modal domain is the same person as the actual interpreter. As for other properties than identity, the idea is to fix the properties of the interpreter outside the scope of the modal operator, while allowing that these fixed properties are not actually instantiated by the actual interpreter. The name of the game is then to take a stance on the audience-properties that characterise the audience that partly characterises the modal domain.[65]

A natural thought, which has been toyed with by several authors, is that the relevant audience-properties are those of a virtuous, or 'ideal' interpreter (King 2014, Neale 2004, Nowak & Michaelson 2021, Strawson 2000). A virtuous interpreter is linguistically and pragmatically competent, alert, attentive, reasonable, etc… The idea is to hold fixed these audience-properties across the modal domain while other audience-properties are allowed to

[64] Thanks to Derek Ball for this point.

[65] Stipulating that the audience across the modal domain is identical with the actual audience might create problems if some other properties of the modal audience are inconsistent with her identity (e.g. being born to different parents). The audience in our range of non-actual possibilities would be defined as the same person as the actual audience, but it would have properties which cannot fail to make her a different person. I ignore these metaphysical difficulties in the main text. Speaks has noticed the potential problem that a modal condition on the interpretation of a non-actual audience can beget *impossible audiences* (Speaks 2016, pp.320-323).

vary. As we have seen, a speaker's communicative act should not be made defective by the mere fact that her audience actually lacks interpretive virtues. So the idea that the audience in our modal domain is a virtuous one is tempting. But this idea is wrong, for at least two reasons. The first reason is the most important one. Speakers craft their message for the specific audience they face. They tailor their utterances to what they think their audience knows, what they think their audience has experienced, etc. I take this to be a commonplace about communicative acts.[66] The quality of a speaker's performance is then a matter of whether the speaker's tailoring of her performance is accurate or not. If there is even a grain of truth to this impressionistic consideration, evaluating what a random virtuous interpreter would understand is beside the point. When a speaker exploits properties specific to her particular audience, a random virtuous interpreter would not understand what she means. But surely this does not blemish the speaker's performance. A *virtuous interpreter* modal condition thus overgenerates defective speaker performances (King 2014, pp.226-229; Nowak & Michaelson 2021, section 3).

The second reason why a *virtuous interpreter* modal condition is not right might be more controversial. As I see it, the audience whose interpretation is relevant to the quality of the speaker's performance is sometimes a non-virtuous one. Say that the interpreter is linguistically incompetent with regard to French, and that this is common knowledge between interlocutors. The speaker has then no business addressing her audience in French: if she does, her communicative performance is defective. But as I understand the *virtuous interpreter* modal condition, the audience across the modal domain does speak French, and thus possibly retrieves the speaker's intended content. Our modal condition incorrectly predicts that the speaker's performance is flawless. The same goes for attention: if it is mutually accepted that the audience is distracted and is not listening to the speaker, then the speaker has no business keeping talking, unless and until she manages to get the attention of her audience back. If she keeps talking to an audience mutually known to be inattentive, her communicative performance is defective. A modal condition based on the interpretation of a virtuous interpreter predicts the opposite: since the relevant audience is attentive, the speaker's performance is faultless. I think this is the wrong prediction.[67]

---

[66] Heck's examples at the end of the previous chapter are nice illustrations of this commonplace. Fricker also gives nice examples of utterances whose intended content would be recoverable by no one except the speaker's audience (Fricker 2012, pp.77-78). These are cases in which the speaker relies for communicative success on mutual knowledge between herself and her hearer that is not shared with anyone else.

[67] What about e.g. a lecturer who keeps talking to a student crowd mutually known to be inattentive (Speaks 2016)? Surely, the lecturer cannot be blamed for producing utterances until the end of the lecture hour, or so the objection goes. That's right, but one should realise that there is more going on with the lecturer than the desire to communicate propositions to her audience. The lecturer is obliged to deliver her lecture as if she had an attentive audience: professional norms or employment contract demand it. There is *extra-communicative pressure* on the speaker to keep producing utterances - imperatives other than having one's intended content recognised. In other cases discussed by Nowak & Michaelson (2021), extra-communicative pressure comes in the form of high practical stakes - e.g. life and death situations. When practical stakes are high, putting up a communicative performance despite knowing that one's audience has properties detrimental to communicative success - e.g. being inattentive - might still be the best course of action overall. The presence of extra-communicative

So much for the virtuous interpreter then. If the audience in our modal condition is neither defined by the properties actually instantiated by the actual audience, nor by virtuous-interpreter properties, by what properties is it defined? Guided by the idea that speakers craft their utterances to what they think their audience is like (including what the audience believes, believes about the speaker, etc), it seems natural to turn to the notion of common ground to characterise the audience that partly defines the modal domain.

One may first suggest that the audience-properties held fixed across the modal domain are those figuring in the common ground of the conversation. Whatever the common ground of the conversation says the audience is like, that's what the audience is like across our modal domain. This is not a novel proposal: King reaches it through roughly the same dialectical route (King 2014). The first difference with King is that he thinks it necessary to make the brute addition that the audience is a virtuous interpreter, no matter what the common ground says the audience is like. I have already argued that this is a mistake. The second difference is that for King the audience not only has the properties the common ground says it has, but also *knows the common ground* (King 2014, p.226). 'Know the common ground' is ambiguous: it could mean *know every common-ground-proposition* (every proposition that is mutually accepted); or it could mean *know of each common-ground-proposition that it is common ground*. It seems to me that King has the second meaning in mind. Overall, King defines the relevant audience in the following manner: (i) it is as the common ground says it is, and (ii) it knows the common ground.[68]

Why should the condition that the audience *know the common ground* be added to the condition that the audience *be as the common ground says it is*? According to Stalnaker, the common ground does not only contain propositions about the topic of the conversation. It also contains propositions about the conversation itself, about what has been done in it, and of course about the participants to the conversation. Hence both a 'first-order' proposition *p* about the topic of the conversation and a 'second-order' proposition that the speaker/audience knows/believes *p* can be part of the common ground. However, and this is where King might be onto something, it seems that a proposition *p* can be part of the common ground without the proposition *that the participants know that p is common ground* being part of the common

---

pressure affects the evaluation of the speaker's action by mitigating our judgement that the action is defective *qua* communicative act. Extra-communicative pressure constitutes extenuating circumstances, and evaluations of the speaker's action sometimes take extenuating circumstances into account. Like other evaluative judgements, judgements on a speech act are all-things-considered judgements (Marsili forthcoming, p.24).

[68] In their (2021), Nowak & Michaelson ignore condition (i) - that the relevant audience is as the common ground says it is - in their evaluation of King's theory. Because of this overlook, they present two cases (*Specialised Skills* and *Shared Memories*) as ones for which King's theory incorrectly predicts something wrong with the speaker's act, although in fact King's theory correctly predicts that the speaker's act is flawless. One reason why it is difficult to discuss their article at this stage is that Nowak and Michaelson focus on the explanation of intuitions about what is said (of pre-theoretical say-judgements) rather than the explanation of intuitions of defective speaker-performances. The connection between intuitions of these two kinds is a delicate issue, as I will explain in the next chapter.

ground. How does this bear on our modal condition for the deficiency of a content-act? Well, if a proposition *p* is part of the common ground but the proposition *that the audience knows that p is common ground* isn't, and if the audience-properties held fixed across the modal domain in our modal condition are just the properties attributed to the audience by the common ground, then the audience across our modal domain does not necessarily know that *p* is common ground. An audience who does not know that *p* is common ground could fail to recognise the speaker's intended content. Our modal condition would then overgenerate defective performances. I take this to be the rationale for King's additional *knowledge of common ground* requirement.

Although the rationale is sound, the additional requirement is not needed. Remember that the modal domain of our modal condition is partly defined by the actual context of the speaker's utterance, and that this context can be construed as the common ground of the conversation prior to the speaker's utterance. Now, at least according to Stalnaker, if *p* is mutually accepted (is common ground), then it is mutually accepted that *p* is mutually accepted (Stalnaker 2002, pp.706-707). In Stalnaker's framework, that a proposition is common ground does entail that every participant takes it to be common ground (by contrast, a participant could take a proposition to be common ground that isn't: this is why the set of presuppositions of a participant can diverge from the common ground). Hence, when we hold fixed the common ground across the modal domain, we also hold fixed that the participants - and *a fortiori* the audience - take common-ground-propositions to be common ground.

I have just argued that King's *knowledge of common ground* requirement on the audience is redundant if Stalnaker's logic of common ground is endorsed. Of course, the knowledge of common ground requirement is redundant only if further conditions are also fulfilled: the common ground is stable across the modal domain, and the audience across the modal domain has the properties attributed to it by the common ground. One might wonder if redundancy cuts both ways: isn't it redundant to require that the audience instantiate the common-ground-properties, when the two other requirements - stable common ground plus knowledge of common ground - are in place? One case from the literature will be enough to dispel this worry:

> Riding the cable car in La Grave, we are presented with an extremely complex high-alpine panorama. Without gesturing, I say: 'Look—that's the line that we skied last year, isn't it', intending to refer to a particular route that weaves through cliff bands and a series of narrow couloirs. You know which line I am talking about, because you were there [...]. (Nowak & Michaelson 2021, p.264)

As Nowak and Michaelson insist, you *had to be there* to know which line the speaker means. Knowing that it is common ground *that the participants skied line X last year* is not enough to know which line the speaker means. Only having actually skied line X will do, because the speaker counts on her audience remembering skiing the line that is now visible from afar. This requires having in fact skied this line.

In summary, I propose that the modal domain over which the 'would' in our modal condition quantifies is defined by (i) the speaker's actual utterance and the communicative intentions underlying it; (ii) the actual context of the utterance (*qua* pre-utterance common ground); and (iii) the audience having the properties that the context attributes to it. This is how I propose to analyse the modal condition that the content-act performed by a speaker is defective just when the audience would not retrieve the speaker's intended content.

3. Stress-testing a common ground modal analysis

Some cases seem to show that my elucidation of the modal condition yields a standard that is too stringent, and some cases that it yields a standard that is too lax. Let me start with the first kind of cases. These are cases in which my analysis of the modal condition seems to predict a defective communicative performance although the performance is intuitively fine. It turns out that these cases have a common structure: the speaker reasonably takes a proposition to be common ground which isn't. She relies on this proposition being common ground in crafting her utterance. Our modal condition, as it has been elucidated, defines the modal domain in terms of what is actually common ground, rather than in terms of what the speaker reasonably takes to be common ground. Our modal condition tends to predict defective speaker-performances in such cases, although intuitively the performance is alright. Or so the worry goes.

Let me present cases that display this structure. Here is my adaptation of a case originally presented by Speaks (Speaks 2016, p.318). Let me call this case *Basketball*. In *Basketball*, a NBA game at Madison Square Garden - home of the New York Knicks - is unfolding. Jalen Brunson, the point-guard and playmaker of the Knicks throws a long no-look pass from a difficult angle to another player, who finishes the move with an easy basket. Someone in the audience tells her neighbour: 'He's amazing, isn't he?'. The speaker is of course talking about Jalen Brunson. She presupposes that her neighbour has enough basketball knowledge to find out who she intends with her use of 'He'. But it is not common ground that the speaker's neighbour has passing basketball knowledge. The speaker's neighbour does not accept that she has passing basketball knowledge, for the simple reason that she does not have passing basketball knowledge. She is at Madison Square Garden watching a Knicks game because someone gave her tickets and attending the game seemed like a fun thing to do. Because of her basketball ignorance, the speaker's neighbour cannot tell whether the speaker is talking about Jalen Brunson or about the player who finished the move with an easy basket.

Let us apply our modal condition to this case. Since it is not common ground that the speaker's neighbour has passing basketball knowledge, the speaker's neighbour does not have passing basketball knowledge in every possibility in the modal domain. Our modal condition thus plausibly predicts that there are some possibilities within this range in which the relevant audience does not retrieve the speaker's intended referent (Jalen Brunson). Hence, our modal condition plausibly predicts that the speaker's performance is defective. Speaks contends that this is the wrong prediction, and motivates his intuitive judgement with

the claim that the speaker's assumptions about her neighbour "*are quite reasonable assumptions to make about someone in the crowd at a basketball game"* (Speaks 2016, p.318).

I agree with Speaks that the speaker's assumption is reasonable to some extent. But the kind of rationale that Speaks gives for giving a pass (no pun intended) to the speaker of *Basketball* does not extend to various other cases of speaker-mistakes about the common ground, including the Carnap-Agnew case. In the Carnap-Agnew case, the speaker believes that Carnap's picture is behind her. That's a reasonable thing to believe: why should the speaker suspect that someone pulled an outrageous prank and switched pictures? The speaker also assumes that her audience can now see Carnap's picture behind her, and she takes it to be common ground that Carnap's picture is behind her. That's a perfectly reasonable assumption. By parity with *Basketball*, we should claim that a false but reasonable assumption that it is common ground that Carnap's picture is on the wall does not make the speaker's performance defective. But the performance *is* intuitively defective! In the face of cases such as *Basketball*, it is tempting to modify our modal condition so that it is not *common-ground-propositions* that define the modal domain, but rather *propositions which the speaker reasonably takes to be common ground*. However, such a move would make the Carnap-Agnew case one of faultless speaker-performance, which I at least take to be an unacceptable result.[69]

Secondly, it is not entirely clear to me that in *Basketball* the speaker's false assumption about the common ground is reasonable. It is not terribly unreasonable, but a basketball game is exactly the type of event for which people hold tickets for other reasons than being a fan or a connoisseur. Now, we can tweak *Basketball* so that the speaker's assumption about the common ground is not just reasonable-ish, but truly unimpeachable. Say that the speaker's neighbour wears a Knicks jersey (a Brunson shirt even!), and cheers when the Knicks score. Suppose that the reality behind this is not that the speaker's neighbour is a Knicks fan, but that she got the jersey for free and is eager to blend in at Madison Square Garden. In this variant of the case, the speaker's assumption that it is common ground that her neighbour has passing basketball knowledge is unassailable. This variant of *Basketball* seems to put more pressure on my analysis of the modal condition. Doesn't my analysis still counterintuitively predict that the speaker's act is defective? Well, maybe not. It seems to me that the neighbour herself should accept for the purpose of small talk with fans around her *that she is a Knicks fan* (and thus that she has passing basketball knowledge). After all, she's wearing the jersey, cheering with the others, etc… She knows that everyone around her will presuppose that she

[69] The 'reasonable expectation' criterion of footnote 61 may now be seen in a different light. The speaker of the Carnap-Agnew case can reasonably expect her audience to recognise her meaning in the following sense: given what she reasonably takes to be her evidence (to be common ground), she can reasonably expect her audience to recognise her meaning. This condition seems to be derivative of the modal condition that the audience *as the speaker reasonably takes it to be* would not get the speaker's meaning. Now, the speaker *cannot* reasonably expect her audience to recognise her meaning in another sense: given the actual evidence (the actual common ground), she cannot reasonably expect her audience to recognise her meaning. This further condition seems to be derivative of my preferred analysis of the modal condition.

is a Knicks fan. There is then rational pressure on her to presuppose, as everyone else does, that she is a Knicks fan, although she knows she isn't one.

Stalnaker has discussed the general case of a participant in a conversation realising that another participant takes *p* to be common ground although *p* is not common ground because she (the first participant) does not presuppose *p*. There are several possible cooperative responses on the part of the first participant, but if this first participant knows that *p* is false and does not find it convenient or necessary to disabuse her interlocutor immediately, the remaining cooperative solution is to go along with her interlocutor and temporarily accept *p* (Stalnaker 2014, pp.47-49). One may call this phenomenon *conciliatory accommodation*. In our variant of *Basketball*, it seems that we are faced with *pre-emptive* conciliatory accommodation: the speaker anticipates that people around her will presuppose that she is a Knicks fan. Instead of waiting for them to make their presupposition manifest, she preemptively makes this presupposition hers. As far as I know, Stalnaker has not discussed the more specific phenomenon of pre-emptive conciliatory accommodation.

Back to the predictions of my preferred modal condition, if in this variant of *Basketball* it is indeed common ground that the speaker's neighbour is a Knicks fan, the interpreter across our modal domain is indeed a Knicks fan (although the actual interpreter isn't). And so my preferred modal condition predicts that the relevant audience would understand that the speaker means Jalen Brunson, and thus predicts that the speaker's content-act is faultless. Speaks puts forward a case called *Sudden Blindness* which displays the same structure as *Basketball*: the speaker reasonably takes it to be common ground that her audience is normally sighted, although her audience has suddenly turned blind (Speaks 2016 pp.315-316). King provides the response I have just provided to the variant of *Basketball*, although he does not use my terminology 'pre-emptive conciliatory accommodation' (King 2013, pp.295-296).

More generally, the idea that a proposition *p* can become or remain part of the common ground although one participant knows that *p* is false can help us deal with various *prima facie* counterexamples to my preferred analysis of the modal condition. I wrote earlier that the fact that the speaker's audience is not a virtuous one - e.g. that she is inattentive - should not automatically make the speaker's content-act defective. Intuitively, the speaker is not responsible for unforeseeable shortcomings of her audience. This is why the *virtuous interpreter* modal condition seems initially plausible. Now, suppose that a speaker's audience becomes inattentive in the course of a conversation, and that the speaker does not immediately realise it. Call this generic case *Sudden Inattention*. The speaker still takes it to be common ground that her audience is attentive, but the audience of course knows that she herself is no longer attentive. Does it follow that the proposition *that the audience is attentive* is no longer part of the common ground, and thus that my analysis of the modal condition incorrectly predicts that the speaker's performance is defective? I would reply that it doesn't follow. A proposition *p* does not cease to be common ground just because it becomes false and one participant knows it has become false. If this participant further knows that other participants are in no position to know that *p* is false, she is under rational pressure to keep

accepting *p*. When a speaker's audience suddenly becomes inattentive in the middle of the conversation, this audience doesn't immediately cease to presuppose that she is attentive. That she is attentive remains common ground, at least temporarily. Therefore, *Sudden Inattention* is not a counterexample to my analysis of the modal condition.

So far I have presented cases in which the speaker is mistaken about the common ground because she herself falsely believes a proposition *p* (and falsely believes that the audience believes *p*, etc). The speaker might alternatively be mistaken about the common ground because *the audience* has a false belief. In yet another variant of *Basketball*, the speaker who utters 'He's amazing' takes it to be common ground that she herself is a basketball fan (she is). Unfortunately, her neighbour does not take the speaker to be a basketball fan. If it is not common ground that the speaker is a basketball fan, it is not clear that our modal condition can predict a faultless performance on the part of the speaker. Is this a problem? I think not. Suppose that the speaker wears no ostensible sign of Knicks fandom, and cheers at unusual things because she has unusual basketball obsessions. It is then reasonable for the audience not to take the speaker to have passing basketball knowledge. But if this is reasonable for the audience, then it is not reasonable for the speaker to take it to be common ground that she (herself) is a basketball fan. And so I have no problem with the modal condition predicting that the speaker's act is defective. Now, make the alternative supposition that the speaker wears a Knicks shirt and cheers at all the right moments. All of this is common ground between herself and her neighbour. It is then quite unreasonable for the neighbour not to accept that the speaker is a Knicks fan. Furthermore, even if the audience is unreasonable in this way, it seems to me that it is common ground *that she is reasonable*. The common ground thus contains the following propositions: that the speaker is wearing a Knicks shirt, that the speaker is cheering at all the right moments, and that the audience is a reasonable person. Given this common ground, my elucidation of the modal condition correctly predicts that the speaker's performance is faultless in this case.[70]

Troubles for my preferred analysis of the modal condition do not end with *Basketball*-type cases. Other cases seem to show that my preferred analysis *undergenerates* defective performances. First of all, the appeal to conciliatory accommodation might have the unintended consequence of undergenerating defective performances. Say that in the Carnap-Agnew case the audience is in on the picture switch. Doesn't the audience then know that the professor presupposes that Carnap's picture is on the wall behind her, and isn't this

[70] In yet another variant of this case, the speaker's neighbour is wearing a Knicks shirt, and so the speaker legitimately takes it to be common ground that her neighbour is a basketball connoisseur. But the speaker's neighbour has impostor syndrome: she doesn't take herself to have passing basketball knowledge although she is quite the expert. It is then not common ground that the audience has passing basketball knowledge. Does our analysis of the modal condition then predict a defective performance in this case? I am not sure. It is common ground that the audience is wearing a Knicks shirt, that the audience is cheering at all the right moments, and also that the audience is reasonable (although she isn't: impostor syndrome is a form of irrationality). An audience defined by these common ground propositions and who knows the common ground would recognise the speaker's intended referent, because this audience would be a basketball fan. My preferred analysis of the modal condition is not threatened by cases of audiences who have unreasonable beliefs about themselves.

reason enough to presuppose with the professor that Carnap's picture is on the wall? If it is common ground that Carnap's picture is on the wall, then my preferred analysis predicts that the speaker's performance is faultless after all.

I am not sure how to think of this variant of the Carnap-Agnew case. The case must be fleshed out in several ways. The audience might accommodate the professor's presupposition that Carnap's picture is on the wall once the professor has uttered 'That's a picture of one of the greatest philosophers…'. This utterance makes the professor's presupposition manifest, and if it is not appropriate to correct the professor, the remaining option is to accommodate the false presupposition. In this scenario, it is the speaker's utterance that triggers conciliatory accommodation. This is not a case of *pre-emptive* conciliatory accommodation. The common ground in which the proposition that Carnap's picture is on the wall figures is the *post-utterance* common ground, and I have argued in the previous section that the common ground defining the modal domain of our modal condition is the *pre-utterance* common ground. Developed as such, this variant of the Carnap-Agnew case involving conciliatory accommodation is not problematic for my analysis.

To make this case really problematic, it needs to be common ground *before the speaker's utterance* that Carnap's picture is on the wall. If the professor shows no sign of noticing the picture switch when entering the classroom, the audience might guess that the professor presupposes that Carnap's picture is on the wall. Is this guess enough to trigger pre-emptive conciliatory accommodation on the part of the audience? I am not sure. In a normal conversational setting, noticing that someone presupposes *p* is a good reason to presuppose *p* with them. But in this case the prankster-audience relishes in the professor's ignorance of the switch, and hence relishes in the professor presupposing false propositions. There is a non-cooperative flavour to this situation, and hence it is not clear to me that the prankster-audience would engage in conciliatory accommodation. Conciliatory accommodation is indeed a *cooperative* tactic.

As Derek Ball pointed out to me, the prankster-audience could gradually adopt a cooperative attitude. Instead of simply laughing when the professor points at Agnew's picture while ranting about Carnap's philosophy and achievements, the prankster-audience could genuinely become interested in Carnap's philosophy and achievements. She would ask questions about Carnap, and would try to remember the professor's answers. If the situation develops in this way, the prankster-audience ends up pretending that Carnap's picture is on the wall in order to gather knowledge about Carnap. This is a cooperative tactic, and one that looks like conciliatory accommodation. The contrast between this further phase of pretence for the sake of learning and the initial phase of relishing in the professor's mistake reinforces my suspicion that the initial phase did not involve accommodation of the professor's presupposition that Carnap's picture is on the wall. Now, it seems that I must admit that in the second, cooperative phase it is common ground that Carnap's picture is on the wall. And thus I am committed to the result that the professor's utterances are faultless in this second phase. I am not sure how bad this is. The professor is in a poor epistemic situation, but it is compensated by the audience's privileged epistemic situation (knowing what the professor

ignores, and that the professor ignores it) and her willingness to be as cooperative as needed to learn from the professor's utterances render the speaker's communicative act non-defective.

Beside cases of conciliatory accommodation, cases that suggest an under-generation of defective performances by my preferred analysis are ones in which a proposition is common ground because both speaker and audience happen to falsely believe it independently. Because the falsely believed proposition is indeed common ground, our modal condition predicts a faultless performance, although we get an intuition of defective performance. Here is a variant of the Carnap-Agnew case that illustrates this structure. Nothing new on the speaker's front: she is not aware of the picture switch, and thus takes it to be common ground that Carnap's picture is on the wall behind her. But check this out: the audience, who is aware that Carnap's picture usually hangs on the wall behind the speaker, has forgotten her glasses today. Unfortunately, she is myopic.[71] She can't see the picture behind the speaker very well, but mistakenly assumes that it is Carnap's picture. In this case, it seems that it is mutually accepted between participants that Carnap's picture is on the wall. Given this common ground, my preferred analysis plausibly yields the prediction that the speaker's performance is faultless. This is problematic to the extent that we still get an intuition of defective communicative performance in this case. I, at least, still get this intuition.[72]

Some of the features of this variant make me doubt that it really is common ground *that Carnap's picture is on the wall*. The audience takes the picture she is now seeing (Agnew's picture) to be the picture she is remembering (Carnap's picture). She thus has a confused *de re* thought. As I explained in section 3.b. of Chapter 1, it is not clear what the content of a confused *de re* thought is. Some authors contend that confused *de re* thoughts are empty (Lawlor 2007, Recanati 2012); some that they partially refer to each of the confused objects (Recanati 2016); and some that they have a non-actual object as their content (Milikan 2000, Unnsteinsson 2019). No extant view has it that only one of the confused objects is the content. But then the false belief of the myopic audience in the problematic case is not that *Carnap's picture* is on the wall. And hence she does not presuppose that Carnap's picture is on the wall. This proposition is not common ground after all.

The unusual feature of confused *de re* thought is specific to the case at hand. It is then fair to ask if the structure I mentioned two paragraphs ago is instantiated by cases which do not

[71] Perry discusses another variant of the Carnap-Agnew case involving short-sightedness (Perry 2009, p.190).

[72] If both speaker and audience notice the picture switch, but together adopt for whatever reason the pretence that Carnap's picture is still on the wall behind the speaker, then it is clear that the speaker's act is faultless when she utters sentences such as 'He would have said that ontology is a risky enterprise' while pointing at Agnew's picture behind her. Since it is common ground that Carnap's picture is still on the wall, our modal condition correctly predicts a faultless performance. Mutual acceptance of a proposition due to a mutually recognised pretence is different from mutual acceptance due to identical but independent false beliefs. Only the latter is potentially problematic for my analysis of the modal condition.

display this feature. Are there cases in which speaker and audience together presuppose *p* because they independently falsely believe *p*, and no confused *de re* thought is involved? I think that there are such cases. In yet another variant of *Basketball*, the speaker's neighbour is wearing a shirt that looks a bit like a Knicks jersey but isn't one. This neighbour is not a Knicks fan: she is just hoping to blend in by wearing a shirt she falsely believes to be a Knicks jersey. The speaker is presbyopic, and mistakenly takes her neighbour's shirt to be a Knicks jersey. Given this setup, we can assume that it is common ground *that the speaker's neighbour is wearing a Knicks jersey*, and that this happens because both speaker and audience independently (falsely) believe this proposition.

Now, what does my preferred modal condition predict in this case? Given the common ground in this case, my preferred modal condition plausibly predicts that the speaker's performance is faultless. Is this a bad prediction? It is reasonable-ish for the speaker to assume that her neighbour is a Knicks fan, regardless of the nature of her shirt, assuming that the neighbour cheers when the Knicks score. Now, if the neighbour does *not* cheer when the Knicks score, the speaker's communicative plan is more reliant on her false perception that her neighbour is wearing a Knicks jersey. But then the speaker, assuming she is reasonable and knows she is presbyopic, should refrain from presupposing that her neighbour is a Knicks fan. So maybe when the speaker's neighbour is not cheering at the right moments it is *not* common ground that she is a Knicks fan. Overall, I am not convinced that mutual presuppositions based on independently false beliefs pose a problem for my preferred elucidation of the modal condition.

I end here the audit of my preferred analysis of the modal condition. There are hundreds more *prima facie* counterexamples out there, and some of these will exert more pressure on the analysis than any of the cases I have presented here. Speaks dismisses this common ground analysis of the modal condition on the basis of cases similar to *Basketball* and *Sudden Inattention* (Speaks 2016, pp.315-319). I have argued that a common ground analysis can absorb the pressure of such cases. Nowak and Michaelson present as counterexamples to a common ground analysis cases in which communicative success depends on the audience having properties which the common ground attributes to it (Nowak & Michaelson 2021). This seems to me strange, as both King's analysis and mine straightforwardly deal with these cases. Peet quickly argues against other analyses of the modal condition, and he assumes that a King-style analysis in terms of common ground is hopeless, citing Nowak and Michaelson (Peet 2021, pp.10-12). I hope to have shown that an analysis in terms of Stalnakerian common ground is more resilient than acknowledged in the literature.

Conclusion

Cases of unintentional divergence between what the speaker means and what she says elicit an intuition of deficient communicative performance. This defect is independent from actual communicative failure, and consists in the speaker's utterance not furthering her communicative aim (section 1). A modal analysis of the defect may be given, according to

which the speaker's performance is defective just when the audience would not retrieve the speaker's intended content. The 'would' in this modal condition quantifies over a modal domain restricted by: (i) the speaker's actual utterance and communicative intentions; (ii) the pre-utterance common ground; and (iii) the audience-properties attributed by the common ground (section 2). Some cases seem to show that this analysis of the modal condition over-generates defective performances, and some cases that it under-generates defective performances. On closer inspection, the analysis can withstand the pressure of these cases (section 3).

Chapter 7: Publicity and responsibility

In section 4 of Chapter 4, I hinted at a theory of the determination of what is said free from the demand that determinants be linguistically encoded. In Chapter 5, I raised the possibility of a unified non-semantic explanation of the publicity of both what is said and what is implied. In this final chapter, I try to fulfil these promises.

In the previous chapter, I explained intuitions of defective communicative performance in cases of unintentional divergence between what the speaker says and what she means. I did so with a modal condition on the interpretation of a common-ground-defined audience. My proposal in this chapter is that the same modal condition can account for our pre-theoretical judgements about the content of speech acts, including what is said. What is said is determined by the interpretation of a common-ground-defined audience. The publicity of both what is said and what is implied (and other intuitive speech act contents) is easily explained if these contents are what a common-ground-defined audience would take the speaker to mean.

In the first two sections of this chapter, I present my view and address some of the difficulties it faces. The next two sections are more tentative and speculative. In the third section, I try to explain why our pre-theoretical judgements track the interpretation of a common-ground-defined audience. I contend that such tracking serves to record the responsibilities and entitlements incurred in making utterances. In the fourth and final section, I discuss possible connections between this contention and two theoretical projects in the philosophy of language: the project of defining assertion, and the project of explaining speaker-deniability.

1. Applying the modal condition to what is said

I claim that the same modal condition that serves to evaluate the quality of the speaker's content-act can serve to predict what a speaker says: the speaker says that *p* if and only if the audience would take her to (primarily) mean that *p*. The range of possibilities expressed by 'would' and the nature of the relevant audience are the same as for the evaluation of a speaker's content-act. The modal domain over which 'would' quantifies is the set of possibilities in which (i) the speaker produces the utterance she actually produces with her actual communicative intentions, (ii) the context of the utterance *qua* common ground is what it actually is, and (iii) the audience has the properties attributed to it by the common ground. This modal condition gives the right result in Carnap-Agnew-style cases. Since the relevant audience would take the speaker to mean something about Agnew's picture, it is correctly predicted that the speaker said something about Agnew's picture. If what a speaker says is what a common-ground-defined interpreter would take the speaker to primarily mean, the *publicity* of what is said is immediately accounted for. What is said is determined by the

interpretation of a common-ground-defined audience, and thus not determined by the speaker's intentions or the audience's actual interpretation.

What a common-ground-defined audience would take the speaker to mean may come apart from both what the speaker means and what her audience actually takes her to mean. Take an example of double linguistic mistake. The speaker utters 'Chloe is evasive' meaning that Chloe is elusive. This is because she falsely believes that 'evasive' means elusive. Her interpreter takes her to mean that Chloe is intrusive. This is because the interpreter falsely believes that 'evasive' means intrusive. Speaker and interpreter ignore each other's linguistic idiosyncrasies, and they take each other to be competent English speakers. Their false beliefs about the meaning of 'evasive' do not make it into the common ground of the conversation, but their being competent English speakers does. Assuming that it is common ground that both speaker and interpreter are acquainted with Chloe, a common-ground-defined audience would take the speaker to mean that *Chloe is evasive*. Hence, our modal condition predicts that the speaker said that Chloe is evasive. This case of double linguistic mistake illustrates how the interpretation of a common-ground-defined audience can come apart from both the speaker's meaning and the actual interpretation of her audience.

Nowak and Michaelson have recently argued that various modal conditions on the interpretation of a non-actual audience yield incorrect predictions about what the speaker said (Nowak & Michaelson 2021). The problematic cases are ones in which the non-actual audience would simply have no idea what the speaker means. In such cases the modal conditions predict that the speaker said *nothing* (or that her content-act was void, or something of the sort). And yet, Nowak and Michaelson claim, we report speakers of such cases as having said what they meant. Given that modal conditions on the interpretation of a non-actual audience are supposed to track the quality of a speaker's communicative performance, the conclusion of the authors is that one can say what one means despite putting up a defective communicative performance.[73]

Nowak and Michaelson do not scrutinise my preferred modal condition. However, it turns out that my preferred modal condition is indeed threatened by the cases that threaten other modal conditions the authors scrutinise. Here are two such cases plucked from the metasemantics for demonstratives literature:

> … I am sitting on Venice beach on a crowded holiday looking south. Hundreds if not thousands of people are in sight. I fix my attention on a woman in the distance and, intending to talk about her and gesturing vaguely to the south, say 'She is athletic.' You, of course, have no idea who I am talking about. (King 2014, p.223)

[73] Nowak and Michaelson put it thus: *"neither of us is convinced that a speaker is actually required to 'do enough' in order for demonstrative reference to succeed"* (Nowak & Michaelson 2021, p.268). 'Doing enough' is a variant of formulations found in the metasemantics for demonstratives literature, such as 'being understandable' and 'putting one's audience in a position to discern one's intention'. The conclusion in the main text is a transposition of Nowak and Michaelson's conclusion outside the metasemantics for demonstratives debate.

> A first responder appears at the scene of a wilderness medical emergency. A second person happens to walk past, sees the first person attending to the victim, and sees an open backpack filled with first-aid supplies. Neither one knows anything about the other, and so the common ground is minimal. The first responder says, 'Hand me that SAM splint' [without pointing]. Fortuitously, the second person knows what a SAM splint is and complies. (Nowak & Michaelson 2021, p.264)

The first case is called *Venice Beach*, the second case *Specialised Vocabulary*. Nowak and Michaelson claim that in *Specialised Vocabulary* the speaker did refer to her intended referent, and thus did request *that her audience hand her the SAM splint* (let us put aside for now the fact that this content is not said but requested). Nowak and Michaelson do not discuss *Venice Beach*, but their general point seems to extend to it. If we really had to ascribe a content to the speaker's act of saying in *Venice Beach*, it would be her intended content - about the woman she means. Yet our modal condition seems to predict that the speaker requested *nothing* in *Specialised Vocabulary*, and said *nothing* in *Venice Beach*. Indeed, a common-ground-defined audience would have no idea what the speaker meant. The common-ground-defined audience would likely respond to the speaker's utterance with a 'Who do you mean?' in *Venice Beach*, and with a 'What item do you mean?' in *Specialised Vocabulary*.

Intuitively, the speaker's performance is defective in both cases. In *Venice Beach*, a vague gesture at a crowd plus an utterance of 'She' does not further communicative success in the absence of a rich common ground. In *Specialised Vocabulary*, using specialised vocabulary to designate an object without pointing at it, when there is no evidence that one's audience masters this vocabulary, is communicatively reckless. These cases then seem to show that a defective content-act does not stop us from judging that the content of this act is still the speaker's intended content. In other words, the deficiency of a content-act is apparently not sufficient to cause a divergence between intended content and reported content. This entails Nowak and Michaelson's conclusion that the speaker can intuitively say what she means although she has not 'done enough' to further communicative success.

I think that Nowak and Michaelson get it half-right. They get it right to the extent that in *Venice Beach* and *Specialised Vocabulary* we do not report the speaker as having said *nothing*. But they are wrong to conclude that this is a damning result for the theory that a speaker says that *p* if and only if the relevant audience would take her to primarily mean that *p*. It is well known that there is a *quotational* use of the verb 'say'. 'Say' can take a linguistic expression between quotes as a complement. Here are just three examples:

> *She said 'Go to hell!'*
> *You say 'sentence', I say 'schmentence'.*
> *- What did she say exactly? - I think she said 'We picked a candidate with more management experience'.*

Of course, philosophers are aware of the quotational use of ‘say’. Some even think that ‘say’ is not just ambiguous between a quotational use and a non-quotational use:

> Unless we give it some special technical meaning, the locution 'what is said' is very far from univocal. It can mean the propositional content, in Stalnaker's sense (horizontal or diagonal). It can mean the exact words. I suspect that it can mean almost anything in between (Lewis 1980, p.97).[74]

To make the point I want to make right now, I only need ‘say’ to have a quotational use and a non-quotational use. ‘Say’ can take a linguistic expression as a complement, or a that-clause. My point is that whenever someone utters a sentence, we do not baldly report her as having said *nothing*. We do not do this because it would pragmatically convey that the speaker remained silent, given the quotational meaning of ‘say’. It is not necessary to provide a piece of Gricean reasoning to show this. It is a simple empirical observation that a bald report ‘She said nothing’ in answer to the question ‘What did S say?’ conveys that S remained silent.

Interestingly, qualified reports seem able to block the *silence* reading of ‘say nothing’ in certain cases. One can report a rambling politician going through stale talking points as follows: ‘He went on and on but said nothing’. In French we even find the idiomatic phrase ‘parler pour ne rien dire’ (literally, *speak to say nothing*), whose target probably overlaps with that of the verb ‘bullshit’ in English. Notice that in such cases, although the implication that the speaker remained silent is obviously blocked, the judgement that the speaker said nothing does not target the speaker’s *content-act*. The politician’s content-acts are not void: a common-ground-defined audience would retrieve the politician’s intended contents. In such a case a ‘say nothing’ report seems instead to target the speaker’s *illocutionary act*: the politician who ‘goes on and on but says nothing’ does not make assertions, or makes defective assertions. Hence, it seems that in ordinary usage there is no room for a meaning of ‘say nothing’ that targets the speaker’s *content-act*, between a meaning that targets *the utterance of a sentence* and a meaning that targets the speaker’s *illocutionary act*.

In *Venice Beach* and *Specialised Vocabulary,* we would not even report the speaker as having said nothing in the context of larger, qualified reports. But do we baldly report the speaker as having said what she means? It seems to me that we don’t. Such reports would sound off without further qualification. Speaking only for myself, in *Venice Beach* I would qualify my report that the speaker said something about the woman she means - call her X - with a statement of the following sort: ‘The speaker meant X, but she didn’t make it clear’. Similarly for *Specialised Vocabulary*, I would report the speaker as having requested the SAM splint, but I would add a negative qualification to the effect that the speaker’s communicative performance was of dubious quality.

[74] See also (Ziff 1972) for the many acceptable pre-theoretical uses of ‘what is said’.

The relationship between my preferred modal condition and pre-theoretical judgements on what is said is then not as straightforward as one might hope. When a common-ground-defined audience would take the speaker to primarily mean something other than what she means - e.g. *q* when the speaker means *p* - we do report the speaker as having said that *q*. But when a common-ground-defined audience would not settle on an intended content, we do not report the speaker as having said *nothing*. We rather tend to report the speaker as having said what she means - that *p* - with further qualifications that emphasise the deficiency of the speaker's performance. I do not think that the correct conclusion to draw from the overall data is that my preferred modal condition is inadequate. For one thing, the ordinary uses of 'say nothing' can help us resist this conclusion. For another, I will argue in the next section that there is evidence that according to our intuitive notion of speech act content the content-act of the speaker is void in *Venice Beach* and *Specialised Vocabulary*, despite our reluctance to report the speaker as having said nothing.

The relationship between my preferred modal condition and say-judgements is not straightforward in yet another way. On my view, what is said is not always what the speaker primarily means, but there is still a constitutive connection between what is said and primary speaker-meaning, since what is said is what a common-ground-defined audience *would take the speaker to primarily mean*. This may seem problematic in view of our ordinary reports of non-literal speech.[75] When someone sarcastically utters 'P' to mean that *q*, don't we report the speaker as having *said that p* though she would be taken to primarily mean that *q*? And when someone metaphorically utters 'P' to mean that *q*, don't we report the speaker as having *said that p* though she would be taken to primarily mean that *q*? The worry here is that say-that judgements on non-literal utterances track *the linguistically determined content of the uttered sentence,* rather than what the speaker would be taken to primarily mean. If this worry is correct, our pre-theoretical notion of *saying that p* is closely related to Austin's notion of a *locutionary act* (Austin 1975). Saying that *p* roughly amounts to intentionally uttering a sentence that semantically expresses *p* (Braun 2011). But then my preferred modal condition misses the mark.

Let me try to dispel this worry. The data on reports of non-literal speech is messy. Empirical studies would be needed, but here I can only provide my own judgements. First, at least some 'say that' indirect reports of metaphorical utterances in which the that-clause targets the *metaphorical* content sound alright to me. To report someone who uttered 'John is a gorilla', clearly meaning in the context of utterance that John is an aggressive alpha male, as having *said that John is an aggressive alpha male* seems fine. Secondly, bald 'say-that' reports of sarcastic utterances in which the that-clause targets *literal* content are not appropriate. Suppose that at my annual review my manager remains curiously silent, before bursting with a sarcastic 'Victor had brilliant ideas this year'. A friend of mine later asks me what was said during my annual review. I cannot reply that my manager *said that I had brilliant ideas this year*. This bald report would not do at all.[76] Camp claims that adding sarcastic intonation to

[75] See (Camp 2006, 2007, 2011) for a thorough investigation of the issues raised by non-literal speech for the semantics-pragmatics interface.
[76] This is also noted by Cappelen and Lepore (1997, p.284).

this report would make it appropriate (Camp 2011, p.4). I am not so sure. This report would be confusing: am I mimicking my manager's original sneering tone, or is the sneering tone mine? My guess here is that the second interpretation is more natural: I would sound as if I was satirising my manager's overenthusiastic praise of my achievements. It would be more appropriate to *mention* my manager's sarcastic intonation - e.g. 'She said sneeringly that I'd had brilliant ideas this year'. Alternatively, a *quotational* (direct) report with sarcastic intonation added would be appropriate. Finally, as Camp concedes, a *qualified* indirect say-report of the literal content of a sarcastic utterance is more appropriate than a bald one (Camp 2007, p.211). For example, when my friend asks me what was said during my annual review, I can appropriately answer: 'My manager said that I had brilliant ideas this year. But she was being sarcastic. She hates my ideas, and hates my guts'.

Where does this leave us? The data on reports of non-literal speech seems to be messy, and the issue would deserve empirical investigation. However, I hypothesise that the idea that the non-quotational use of 'say' exclusively targets semantic/locutionary content will not withstand empirical scrutiny. It is possible that 'say' is not just two-way ambiguous between one quotational meaning and one non-quotational meaning over which we must fight (semantic content, or primary speech act content?). 'Say' might be three-way ambiguous between a quotational meaning, a semantic content meaning, and a primary speech act content meaning. This would explain why on the one hand it is not appropriate to report my sarcastic manager as having said that I had brilliant ideas this year, and on the other hand why sarcastic speakers *mean the opposite of what they say*, as the commonplace has it. The examples provided by Cappelen and Lepore (1997) suggest that three-way ambiguity does not begin to capture all that 'say' allows.[77]

I cannot discuss this difficult issue any further here. The weak conclusion I want to draw is that 'say' has a non-quotational meaning such that the that-clause following 'say' targets the content of a speech act of a type which allows propositional content (stating, supposing, guessing etc). Nothing stops me from allowing that 'say' has a further non-quotational meaning on which the that-clause targets *semantic/locutionary content*. I just need to qualify my contention that what a speaker says is what she would be taken to primarily mean, by specifying that this is a theory of only one meaning of 'say'. I am assuming that we can pre-theoretically distinguish between the various non-quotational meanings of 'say' and isolate one meaning to theorise about. It seems to me that these methodological assumptions are shared by Camp, who takes 'say' to have the two non-quotational meanings I have discussed (Camp 2007, 2011).

[77] According to these authors, some acceptable ordinary 'say-that' reports even target *secondary* contents (e.g. what is implied). Cappelen and Lepore emphasise the extreme liberality of say-that reports to make the point that say-that reports cannot be used as evidence to build a semantic theory. They further think it illegitimate to appeal to 'refined intuitions' about what is said, because this would amount to introducing a theoretical notion of *what is said* that aligns with one's favourite semantic theory. I am not sure that their conclusion is warranted, but in any case I am not claiming anything about the connection between pre-theoretical say-that judgements and semantic theory in the present chapter.

2. Applying the modal condition to other speech act contents

My presentation of *Specialised Vocabulary*, in which the speaker does not *say* something but rather *requests* something, is not innocent. I take it that what a speaker asks or requests is also a matter of what a common-ground-defined audience would take the speaker to mean.[78] I thus want to extend the use of my preferred modal condition from acts of saying to content-acts of other types. This extension raises a difficulty for the notion of a content-act. With the use of verbs such as 'say', 'request', and 'ask', don't I smuggle in *illocutionary forces* in my supposedly force-free content-acts? One possible response is that these verbs reflect *grammatical mood* rather than illocutionary force: namely the declarative, imperative, and interrogative moods. There is a consensus in the literature that mood is distinct from illocutionary force, because grammatical mood underdetermines illocutionary force (Green 2020, Roberts 2018). I could then clarify my notion of a content-act by specifying that a content-act includes *mood* in addition to content. The verbs I invoke to describe content-acts ('say', 'ask', 'request'…) would then be taken as markers of mood. We would then get as many types of content-acts as there are moods: declarative, imperative, and interrogative.

This response is not the right one for me though. I argued in Chapter 5 that what speakers *imply* (suggest, hint, insinuate) is also public. And I suggested that a unified explanation of the publicity of both what is said and what is implied should be sought. I wish to provide such a unified explanation, by extending the use of my modal condition from content-acts associated with *primary* speaker-meaning to content-acts associated with *secondary* speaker-meaning. This theoretical goal prevents me from defining types of content-acts in terms of grammatical mood. This is simply because the type of a 'secondary' content-act performed by a speaker - e.g. implying a proposition - does not necessarily correspond to the grammatical mood of the sentence she utters. For example, a speaker can ask something by saying something with the utterance of a declarative sentence (Braun 2011).

It seems to me that I should define types of content-acts in terms of *semantic type*, i.e. the type of content-entity involved in a content-act. Verbs such as 'say', 'request' and 'ask' in my description of content-acts capture the semantic type of the content involved. On this view, asking something is a matter of putting forward a content of the *propositional-set* type, according to a mainstream semantic treatment of questions (Hamblin 1973). As for the verbs 'say' and 'imply', they are appropriate for content-acts of the propositional type. 'Say' and 'imply' capture more than the propositional type of the content involved though. 'Say' indicates that the propositional content-act in question is associated with *primary* speaker-meaning, while 'imply' indicates that the propositional content-act performed is associated with *secondary* speaker-meaning. It is then possible to categorise content-acts in terms of both (i) the semantic type of the content involved and (ii) whether the content-act is

[78] This sentence might suggest that *illocutionary force* itself is determined by the interpretation ('uptake') of a common-ground-defined audience, but here I am using verbs such as 'ask' and 'request' to describe content-acts, rather than illocutionary acts. I will return to the issue of the determination of illocutionary force in section 4.

associated with primary or secondary speaker-meaning. This yields six types of content-acts, assuming that there are three semantic types.

Let me turn to the explanation of the publicity of what is implied, which is the main point of the present section. I claim that my modal condition not only accounts for what is said but also for what is implied. What a speaker implies is what a common-ground-defined audience would take the speaker to secondarily mean. The distinction between primary and secondary content-acts can thus be understood in terms of primary and secondary speaker-meaning. As a reminder, a speaker-meant content *c2* is secondary when the speaker intends her audience to retrieve *c2 by retrieving another content c1*. To be clear, the distinction between primary and secondary content-acts is not drawn in terms of speaker-meaning in the 'normal' way. The claim is not that the content of a secondary act is what *the speaker* secondarily means. The content of a secondary content-act is rather what *a common-ground-defined audience* would take the speaker to secondarily mean.

Let me illustrate how my preferred modal condition predicts what speakers imply with cases pulled from Chapter 5. In *Albert and Bettina*, Bettina utters 'There's aspirin in the bathroom cabinet' in response to Albert's complaining about his headache. Bettina has outlandish beliefs about headaches, aliens, and aspirins. Let me choose one version of this case in which Bettina is an otherwise rational person, who would not produce an utterance if she did not think that her audience could recognise the communicative intentions underlying it. Bettina takes it to be common ground that she has these beliefs linking headaches to aliens and aliens to aspirin. She (secondarily) means that Albert should grab a tablet of aspirin, hold it against his breast and run as fast as he can. Maybe the secondary speaker-meant content could be seen as being of the imperative type: Bettina commands Albert to take a tablet, to hold it against his breast, and to run. In any case, Bettina believes that Albert is in a position to retrieve this intended content, because she takes it to be common ground that she has beliefs linking headaches to aliens to aspirins. Unfortunately, Albert is not aware of Bettina's outlandish beliefs. It is then *not* common ground that Bettina has these beliefs. Given the actual common ground between Albert and Bettina, a common-ground-defined audience would not recognise Bettina's meaning. This audience would take Bettina to mean that ingesting an aspirin will soothe Albert's headache. And so the modal condition correctly predicts that Bettina implied that ingesting an aspirin will soothe Albert's headache.

In *Alf and Barb,* Alf tells his friend Barb: '*Barry Lyndon* is showing at the cinema tonight. Wanna join?'. Barb answers: 'My sister's in town!'. Barb's sister is a huge Kubrick fan who would never miss a screening of *Barry Lyndon*. Barb mistakenly takes this to be common ground: she misremembers telling Alf about her sister's love of Stanley Kubrick. The common ground before Barb's answer unfortunately does not include this piece of information. Barb means that she will join with her sister, but an audience defined by the common ground would not recognise what Barb means. Such an audience would likely take Barb to mean *that she cannot join*. Our modal condition thus correctly predicts that Barb implied that she cannot join.

In these two cases the speaker makes a mistake about the common ground. The speaker relies on a proposition she takes to be common ground to convey her intended message, but this proposition is in fact not common ground. Such cases yield a divergence between what the speaker secondarily means and what she implies, and our modal condition can easily account for this. There are also variants of these cases in which it is the hearer who makes a mistake by taking a proposition that is not common ground to be common ground. The hearer takes the speaker to be relying on this proposition, and thus infers that the speaker means something other than what she actually means. For example, there is a variant of *Albert and Bettina* in which *Albert* has outlandish beliefs about aliens and aspirins and mistakenly takes these beliefs to be common ground. Bettina means that Albert can soothe his headache by taking an aspirin from the cabinet. Albert takes Bettina to mean that he should hold the aspirin against his chest and run. In this case, our modal condition predicts that Bettina's secondary content-act is faultless, and that Bettina implied what she meant - that Albert can soothe his headache by ingesting an aspirin.

In Chapter 5, I presented a variant of *Alf and Barb* in which both Alf and Barb are mistaken about the common ground. Barb relies on it being common ground that her sister is a Kubrick fan. She means that she will join with her sister. Alf takes it to be common ground that Barb's sister is a chaotic person, and takes Barb to rely on this being common ground. So Alf takes Barb to mean that she cannot commit to a plan for tonight. Both are mistaken: it is not mutually accepted that Barb's sister is a Kubrick fan, and it is not mutually accepted that Barb's sister is a chaotic person. Given the actual common ground of their conversation, a common-ground-defined audience would likely take Barb to mean that she cannot join. This case is supposed to be one of double divergence of what is implied from what the speaker means and from what her audience takes her to mean. Our modal condition is able to account for this double divergence.

Still in Chapter 5, I argued that if there is a unified explanation of the publicity of both *what is said* and *what is implied*, then one should favour this explanation. To a certain extent, I have now provided such an explanation: intuitive speech act contents are public because they are determined by what a common-ground-defined audience would take the speaker to mean. However, there is a sense in which a semantic explanation of the publicity of what is said gives us more. It provides a natural account of our pre-theoretical say-judgements. What do we do when we judge what people say? According to the semantic explanation, we track the content determined by the meaning of the uttered sentence. My modal condition does not seem to provide this layer of explanation. The following question seems legitimate: why do we track what a common-ground-defined audience would take the speaker to mean? I sketch an answer to this question in the next section.

3. Speech-reporting as responsibility-tracking

Speech, even of the most innocuous sort, has consequences. When you tell me that the next train to Versailles leaves at 3.17pm, and I decide to believe you, I act accordingly. A

communicative act has the power of shaping its audience's mental states and actions. Let us take the example of assertoric speech. If I take you to primarily mean that *p* with your utterance, further take your utterance to bear assertoric force, and further take you to be sincere, I will usually take you to believe *p* and further consider whether to believe *p* myself. In Stalnakerian terms, I will consider whether to accept *p* and thus update the common ground of our conversation (Stalnaker 1999).

The claim that speech has consequences on its recipients is commonplace, but it is also too weak. Linguistics acts of communication do not just happen to produce mental effects in their recipients. Rather, the very point of such acts is to produce these mental effects. That a mental effect on the audience obtains is not a by-product of intentional speech, but what one intends when one intentionally speaks. Now, because speech can change other people's mental landscape and comes with the intention to do so, speech comes with responsibility. A speaker is responsible for the mental effects of her speech on others. But what the speaker is responsible for is not the mental effect her speech *actually* has on her audience. It takes two to communicate. The audience, because she must exercise her interpretive capacities on the speaker's utterance, is also partly responsible for the mental effect produced by the speaker's utterance. In line with the previous sections of this chapter, I think that the responsibility borne by the speaker can be identified with the mental effect that *would* be produced in a common-ground-defined audience.

An utterance typically has multiple mental effects between which an explanatory relation holds. When I tell you 'The next train leaves at 3.17pm', you take me to intend for you to believe that I believe that the next train leaves at 3.17pm. If all goes well, you take me to believe that the next train leaves at 3.17pm. And if all goes well, you come to believe that the next train leaves at 3.17pm. The third effect obtains (partly) because the second effect obtains. And the second effect obtains (partly) because the first effect obtains. Within this chain of mental effects, the effect that matters for present purposes is the one that counts as *understanding the utterance* - 'uptake' in speech-act-theoretic parlance. One can call this effect of an utterance its *illocutionary effect*. Take our train example. You can understand what I'm doing when I utter 'The next train leaves at 3.17pm' without forming the belief that *the next train leaves at 3.17pm*. After all, you might suspect that I am ill-informed. So the illocutionary effect is not that the audience believes a proposition. Beyond this negative remark, I will remain neutral on the illocutionary effect of assertion in what follows, sometimes using a toy view for the sake of illustration.

Whether or not speakers are responsible for other mental effects of their utterance on their audience, I am here focusing on the idea that they are responsible for its illocutionary effect (on a common-ground-defined audience). And one may call the responsibility for an illocutionary effect the *illocutionary responsibility*. Now, just as one can abstract a content-act from an illocutionary act, one can abstract a responsibility for the effect of a content-act from an illocutionary responsibility. A speaker who performs a content-act is responsible for the content that a common-ground-defined audience would take her to put forward. So here is my idea of what we do when we report the content of speech acts: we report *the content the*

speaker is responsible for. This content is the content embedded in the illocutionary effect that the utterance would have on a common-ground-defined audience. We thus get an additional layer of explanation. Our judgements on the content of utterances track what a common-ground-defined audience would take the speaker to mean because this content is the content the speaker is responsible for, and ultimately in reporting utterances we track the illocutionary responsibilities incurred in producing utterances.[79]

One can look at things from the perspective of the audience rather than that of the speaker. What the speaker is responsible for is what her audience *is entitled to*. In the assertoric case, assuming for the sake of illustration that the illocutionary effect is for the audience to take the speaker to believe that *p*, one may claim that the audience is entitled to take the speaker to believe that *p* just when a common-ground-defined audience would take the speaker to believe that *p*. And just as one can abstract the responsibility that comes with a content-act from the responsibility that comes with the encompassing illocutionary act, one can abstract the entitlement that comes with a content-act from the entitlement that comes with the encompassing illocutionary act. The speaker's audience is entitled to a content *c* just when a common-ground-defined audience would take the speaker to mean *c*. This idea fits nicely with the point that the speaker is not always responsible for what her audience actually takes her to mean. Sometimes the audience takes the speaker to mean that *p* but a common-ground-defined audience would take the speaker to mean that *q*. In such cases the audience is entitled to *q* but not to *p*.

Speaker-responsibility comes with a derivative speaker-entitlement. If a speaker is responsible for the mental effect her utterance would have on a common-ground-defined audience, she is also *entitled to expect* this effect on her audience. Symmetrically, a hearer's entitlement to a mental effect comes with a derivative responsibility. If a hearer is entitled to the effect the speaker's utterance would have on a common-ground-defined audience, then this hearer is also *responsible for the speaker's expectation* of this very effect on herself (on the hearer). So the speaker is also entitled to something, and the hearer is also responsible for something. Speaker-entitlement and hearer-responsibility may be seen as derivative from speaker-responsibility and hearer-entitlement. Speaker-entitlement and hearer-responsibility are thus the reverse side of the coin.

I wrote earlier that there is evidence that content-acts in *Venice Beach* and *Specialised Vocabulary* are void, despite the reluctance to report the speaker as having *said nothing* in these cases. Intuitively, in *Venice Beach* the speaker is not entitled to expect her audience to take her to mean something about her intended referent. What the speaker is actually entitled to expect is that her audience will not know who she is talking about. Correspondingly, the audience is not responsible for the speaker's potential expectation that she (the audience) will

[79] Korta and Perry (2010, pp.115-120) acknowledge that the (or rather one) pre-theoretical notion of *what a speaker says* tracks what the speaker is responsible for, but choose to introduce a slightly different theoretical notion (for which they use the terminology 'locutionary content'). I do not know whether Korta and Perry think that the pre-theoretical notion of what is said cannot be theorised about, but in any case that notion is not useful for their theoretical purposes.

take her to mean something about her intended referent. The same can be said of *Specialised Vocabulary*: the speaker is not entitled to expect her audience to recognise which tool she means; and the audience is not responsible for the speaker's expectation that she recognises which tool she means.

It is then the reverse side of the responsibility-entitlement coin (speaker-entitlement and hearer-responsibility) that seems best placed to illuminate cases in which a common-ground-defined audience would not settle on a definite interpretation of the speaker's utterance. The obverse side of the coin (speaker-responsibility and hearer-entitlement) seems more useful in accounting for cases in which a common-ground-defined audience would take the speaker to mean something other than what she means. In the Carnap-Agnew case, the speaker seems responsible for her audience taking her to mean something about Agnew's picture, and the audience is entitled to taking the speaker to mean something about Agnew's picture.

My proposal is then that our practice of giving verdicts on the content of speech acts is that of tracking the content embedded in the illocutionary responsibilities and entitlements incurred in producing utterances. One can think of this practice as 'conversational refereeing': we adjudicate which responsibilities and entitlements have been incurred by the speech acts performed in a conversation. Participants might have their opinion on which responsibilities and entitlements have been incurred. But the opinions of participants do not prevail, either directly with one participant's opinion having authority, or indirectly by mutualising the responsibilities and entitlements which all participants agree have been incurred.[80]

The sports metaphor I used in the previous paragraph might be misleading, because my view is inconsistent with several developments of the notion of *conversational score*. This notion was made famous by David Lewis (Lewis 1979). Lewis considers the idea that the score of a conversation is determined by laws which say how the score evolves as a function of the state of the conversation and of the moves made by participants. 'Laws' in the previous sentence is not to be understood in the empirical generalisation sense - e.g. 'laws of physics' - but in the literal sense of conventions associated with an institutionalised practice. My view does not seem to mesh with this idea. When we referee a conversation, we do not follow a rulebook. Neither linguistic nor extra-linguistic rules explain our tracking of the interpretation of a common-ground-defined audience. More generally, reporting speech acts is not a practice governed by conventions.

---

[80] (Geurts 2019) puts forward a normative analogue of the notion of common ground according to which the 'common ground' contains mutual commitments: A and B are committed in a certain way to *p*, committed to each other being committed in this way to *p*, etc… Geurts insists that commitment is not a psychological notion, which means that A and B can be mutually committed to *p* without knowing/believing that they are so committed (Geurts 2019, p.4). If this is right, Geurts' view may be seen as a fully non-mentalistic, yet non-conventionalist conception of the conversational score. On the other hand, Geurts speaks of a commitment to being committed to *p* as *acceptance that* one is committed to *p*, and he holds that A being committed to *p* entails all participants accepting that A is so committed (Geurts 2019, p.19). It seems then that Geurts does not conceive of *acceptance* as a mental state, which is not exceptional in the literature - see (Kölbel 2011) for example.

The second idea considered by Lewis is that *"conversational score might be operationally defined in terms of the mental scoreboards—some suitable attitudes—of the parties to the conversation"* (Lewis 1979, p.346). Stalnaker's notion of common ground may be seen as exemplifying this conception of the conversational score, since the common ground is the set of propositions mutually accepted by participants in a conversation (Stalnaker 2014). Does my view commit me to such a notion of conversational score? In some sense, no: the responsibilities and entitlements incurred by utterances are not determined by the actual mental states of the participants. But in another sense, yes: the responsibilities and entitlements are determined by mental states, although not the actual mental states of the participants.

One aspect of my view is that the content a speaker is responsible for can come apart from what enters the common ground following her utterance. For example, in cases of double divergence, the speaker's utterance might result in no addition of a proposition to the common ground. The speaker intends *p* to be added to the common ground, but her audience takes *q* to be intended to be added. Assuming that the audience is willing to add *q*, the net result of the utterance is that the speaker adds *p* to her own presuppositions and the audience adds *q* to her own presuppositions. No proposition - neither *p*, nor *q*, nor some other proposition - is added to the common ground. By contrast, if a common-ground-defined audience would take the speaker to mean that *r*, then the speaker is responsible for *r*.

Divergence between responsibility and common ground can go the other way around: the speaker's intended content is added to the common ground but she is not responsible for this content. Here is a variant of *Venice Beach* that illustrates this possibility.[81] The speaker is suspected by intelligence agencies of selling state secrets. Unbeknownst to the speaker, her audience is an undercover intelligence operative who has gained her trust over several weeks. The speaker utters 'She's athletic' thinking that her pointing gesture is precise enough, although it isn't. Unbeknownst to the speaker, the intelligence operative knows an awful lot about her. She knows that the speaker has a soft spot for red-haired women with broad shoulders. And so when the speaker utters 'She's athletic', the intelligence officer looks in the (vague) direction of the pointing gesture and immediately recognises who the speaker means: the only red-haired woman with broad shoulders in this direction. In this case, the intrusive knowledge of the audience about the speaker is not common ground. A common-ground-defined audience would not recognise who the speaker meant, and so the speaker is responsible for no content. But the undercover agent does recognise who the speaker means, and the speaker's intended content gets added to the common ground. Overall, the set of responsibilities incurred in a conversation can diverge from the common ground, but the former does not float free from the latter, since for any given utterance the incurred responsibility is determined by the common ground.

[81] The example is inspired by King (2013, fn 7, p.307).

Beyond the *conventional* conception and the *common ground* conception of conversational score, it seems to me that my view is hard to fit with any conception of conversational score, given certain desiderata on the notion of conversational score. If the score after an utterance must be a function of the score before this utterance, then my view is not a version of the view that our pre-theoretical judgements track the conversational score. Here is why. My view is that the content the speaker is responsible for with an utterance U_1 is determined by the common ground of the conversation before U_1. But we have just seen that the content the speaker is responsible for may be distinct from the content that enters the common ground (as in the case of the all-knowing undercover officer). When this happens, and when the next utterance U_2 is produced, the responsibility incurred in making U_2 is determined by the common ground after U_1, rather than by the responsibility incurred in making U_1. In the case of the undercover officer, no content-responsibility is incurred with the utterance of 'She's athletic', but the proposition meant by the speaker enters the common ground. When the next utterance is produced, the responsibility incurred by this further utterance is determined by the common ground that contains the proposition meant by the speaker with her utterance of 'She's athletic'. Hence, in my view the 'score' at a given point in the conversation is not a function of the 'score' before this point.

Because of this feature, I doubt that my view can be taken as a version of the view that in reporting speech acts we track the conversational score. Of course, one could stipulate that the evolving common ground is part of the score, and thus that the responsibility incurred by an utterance U is a function of an aspect of the score prior to U. But I take it that if the responsibility incurred by an utterance is not a function of the responsibilities incurred by previous utterances, then the 'score' metaphor is not apt.[82] Camp introduces the terminology of 'conversational record' and distinguishes conversational record from conversational score, for reasons I cannot examine here (Camp 2018, pp.59-60). I am happy to adopt her terminology for now: the conversational record is the set of responsibilities and entitlements incurred in a conversation.

4. Points of contact: defining assertion and explaining deniability

Here is a picture I like. By tracking the interpretation of a common-ground-defined audience, one tracks the responsibilities and entitlements incurred by utterances. This picture comes into contact with other projects pursued by philosophers of language. First, there is the project of *defining assertion*. This project consists in individuating assertoric force among the illocutionary forces. It amounts to characterising what is common to all and only assertoric acts among illocutionary acts (Goldberg 2015, MacFarlane 2011). One family of theories of assertion defines assertion in terms of *commitment* to the truth of a proposition, and specifies what this commitment consists in exactly (Brandom 1983, Shapiro 2020, Marsili forthcoming).

[82] Thanks to Derek Ball for discussion of this point.

How does this view relate to my claim that the content of an utterance is the content a speaker makes herself responsible for? According to the picture I like, illocutionary acts of all types incur a responsibility *for the illocutionary effect such acts would have on a common-ground-defined audience*. This picture is consistent with individuating assertoric force - and illocutionary forces in general - in terms of *further* speaker-responsibilities. I could endorse the view that assertoric acts are unique among illocutionary acts in incurring further responsibilities X, Y, Z in speakers, beyond responsibility for the illocutionary mental effect produced on a common-ground-defined audience. Such a view of assertion can be part of a more general account of illocutionary force in terms of responsibilities and entitlements: each illocutionary act-type is individuated in terms of the specific responsibilities and entitlements incurred beyond those pertaining to the illocutionary mental effect of the act. Such *normative effects* accounts of illocutionary force have their proponents (Gazdar 1981, Geurts 2019, Goldberg 2023, Kukkla & Lance 2009, Sbisà 1984).

The picture I like is also consistent with a general theory of illocutionary force in which each illocutionary type is defined in terms of the specific mental effect which speakers make themselves responsible for by performing an act of this type. Different illocutionary acts have different mental effects, and thus incur different responsibilities. Let us take assertion as an illustration. The view under consideration is that S asserts that *p* if and only if a common-ground-defined audience would [insert the mental effect that individuates assertoric force, with proposition *p* embedded]. Assuming for the sake of illustration that the mental effect which individuates assertion is *recognising that the speaker intends the audience to believe a proposition*, we get: S asserts that *p* if and only if a common-ground-defined audience would *take S to intend them to believe p*. This view would be the most economical one for me to take: I claimed earlier that the responsibility incurred by a content-act is abstracted from the responsibility incurred by the encompassing illocutionary act. And I further claimed that the illocutionary responsibility in question is that for the mental effect the illocutionary act would have on a common-ground-defined audience. Individuating illocutionary types (and *a fortiori* assertion) in terms of type of mental effect would allow me to exploit the theoretical resources already at my disposal, without digging for additional ones.

One problem that an individuation of assertoric force in terms of mental effect would face is common to several individuations of assertoric force (MacFarlane 2011, Marsili forthcoming, Peet 2021, Stainton 2016). How does one distinguish typical-asserting from the illocutionary act performed in implying something? So far I have used ‘imply’ to designate a secondary *content-act*. Assuming that one wants to treat this content-act as part of an illocutionary act, definitions of assertion struggle not to count implyings as assertings. Don’t speakers also commit to the truth of what they imply? Don’t they express their belief that *p* when they imply that *p*? Isn’t the mental effect on an audience of implying something the same as that in central cases of asserting? To deal with this challenge, one option for a mental effect account of assertion is to explicitly appeal to the primary/secondary speaker-meaning distinction in the definition of assertion: S asserts that *p* if and only if a common-ground-audience would take the speaker to *primarily* mean that p and would e.g. recognise that the speaker intends

her to believe that *p*. Another, less conciliatory option, is to do nothing about the definition of assertion and let the primary/secondary distinction do its work outside the definition. In central cases of assertion (the ones that some philosophers think are inconsistent with implying), the content-act niched in the illocutionary act of assertion is *primary*. One can then take a content-act of implying as part of an illocutionary act of asserting, with the crucial qualification that the embedded content-act is *secondary*. It seems to me indifferent whether the pre-theoretical difference between central cases of asserting and cases of implying is accounted for in the definition of assertion or outside of it.[83] One advantage of counting the illocutionary act that contains a content-act of implying as a species of assertion is that the promise of individuating assertion in terms of mental effect is fulfilled. By contrast, if something more than mental effect is invoked in the definition of assertion, one must give up the idea that a certain mental effect is uniquely associated with assertoric force.[84] [85]

If I endorsed a biconditional of the form 'S asserts that *p* if and only if a common-ground-defined audience would [insert the mental effect that individuates assertoric force, with proposition *p* embedded]', I would take a stance on a question distinct from the question of the individuation of assertoric force. This distinct question is: which facts

[83] García-Carpintero (2018) supports the move of treating the illocutionary act associated with implying as a species of assertion (*indirect* assertion). Marsili writes that the decision to count implyings as assertions is *"primarily terminological"* (Marsili forthcoming, footnote 11). I take him to mean that for most purposes it does not matter whether we account for the pre-theoretical difference between central cases of assertion and cases involving implying within a definition of assertion, or outside of it by positing different species of assertions.

[84] Most *commitment* theorists of assertion yield on this front. They recognise that there is no commitment uniquely associated with central cases of asserting, and they appeal to an extra-ingredient to rule out cases of implying. This extra-ingredient is often that the content of an assertion is *said*, or is the semantic content of the uttered sentence (Marsili forthcoming, Shapiro 2020). Viebahn's work suggests that this retreat may be resisted (Viebahn 2017, 2021). Viebahn's idea is that asserting that *p* is uniquely associated with a commitment to defend that one *knows* that p, which is supposed to rule out at least some acts of implying. Another way to keep a commitment account of assertion 'pure' might be to claim that the commitment associated with assertion comes with an absence of *deniability* (Fricker 2012).

[85] Another issue is that of distinguishing asserting from weaker illocutionary counterparts encompassing *primary* content-acts. Here I have in mind illocutionary acts such as *conjecturing* and *guessing*. Commitment accounts can use the idea that a speaker who conjectures something makes herself open to weaker challenges than a speaker who asserts (Green 2013, Viebahn 2021). It is not clear how a mental effect account can handle this issue. Here is a very tentative suggestion: if we switch from belief to credence, and for example take the essential mental effect to be that the audience recognises that the speaker intends the audience to take the speaker to [insert epistemic state] *p*, we might associate assertion with high credence and its weaker counterparts with low credence.

determine that a speech act with force X has been performed?[86] This question is the illocutionary analogue of the question that has occupied me for much of this dissertation, i.e. which facts determine the content of a given speech act. Is it linguistic conventions? Is it the speaker's intentions ? Is it the audience's interpretation? I have suggested that intuitive content is determined by the interpretation of a common-ground-defined audience. The above biconditional commits me to an analogous answer to the illocutionary analogue of the content-determination question. This answer is: the illocutionary force of a given speech act is determined by the interpretation of a common-ground-defined audience.[87]

It is interesting to note that the possible determinants of illocutionary force largely overlap with those of content. Is it linguistic convention that determines the operative force? Is it the speaker's intention? The audience's interpretation ('uptake')? Also interesting is the fact that when it comes to illocutionary force, answers to the individuation question are to some extent independent from answers to the determination question. For example, whether one takes assertoric force to be individuated by normative effects or by mental effects, one can ask: is it the fact that a speaker intends to commit to *p* which determines that the speaker asserts that *p*, or is it the fact that the audience takes the speaker to commit to *p*? Is it the fact that the speaker intends her audience to recognise that she intends her to believe *p* which determines that the speaker asserts that *p*, or is it the fact that the audience recognises this?

Let me turn to the relation between my views and the project of explaining speaker-deniability.[88] I will focus on Camp's view, as it provides the most fruitful contrast with my views (Camp 2018). Camp explores the phenomenon of *deniability* through the phenomenon of *insinuation*. Camp takes it that to insinuate something is roughly to imply something with deniability. According to Camp, what is *prima facie* puzzling about insinuation, and about deniability more generally, is first that the speaker may genuinely mean what she insinuates. And for this to happen, the speaker must not only intend her audience to

[86] What I call the determinants of an illocutionary act are what Bach and Harnish call the 'success conditions' of a speech act, as opposed to its 'felicity conditions' (Bach & Harnish 1979). 'Success/felicity conditions' seem to map onto Austin's 'A and B/delta-rules' and onto Searle's 'constitutive/regulative rules' (Austin 1975, Searle 1969). As Austin explained, the two questions 'Is act X performed?' and 'Is act X felicitous performed?' can be distinguished. Cases in which a negative answer is given to the first question are cases of *misfire*; cases in which a negative answer is given to the second question are cases of *abuse* (Green 2020). A classic case of misfire is someone attempting to baptise a ship without the authority to do so. A classic case of abuse is lying, which can be roughly understood as insincere assertion.

[87] The distinction between the determination question and the individuation question is largely recognised by speech-act theorists. Goldberg calls the determination question the 'implementation question', and distinguishes it from the individuation question, which he calls the 'characterisation question' (Goldberg 2023). To give just one more example, Green wants to individuate illocutionary types in terms of the specific commitments that come with performances of acts of each type, but he also aims to answer the further question of whether intending to undertake a commitment uniquely associated with illocutionary act of type X is necessary or sufficient to perform an act of type X (Green 2020).

[88] I will not use the phrase 'plausible deniability', because the deniability discussed here is often not plausible at all. This has been noted by most authors in this literature (Camp 2018, Dinges & Zakkou 2023, Lee & Pinker 2010).

entertain the insinuated content, but intend it *overtly*: she must intend her audience to recognise that she has this intention. The second puzzling feature of insinuation and deniability is that a speaker may enjoy deniability even in cases where the insinuated content is successfully communicated, and thus gets added to the common ground. How can one be able to deny that one meant that *p* when one is recognised by one's audience to have meant that *p*, and when *p* becomes mutually accepted by participants in the conversation?

According to Camp, the answer is that the actual context in which an utterance is made is not relevant to deniability. If it was, then when neither speaker nor hearer is mistaken about the common ground and a content is successfully communicated, the speaker would not enjoy deniability. But insinuation shows that speakers do enjoy deniability in many such cases. Deniability over a speaker-meant content *c* is rather a matter of quantification over non-actual contexts.[89] If there is *some plausible context* in which the speaker does not mean *c*, then the speaker enjoys deniability over the content *c* she actually means. The idea of a plausible context is elucidated by Camp in terms of a plausible common ground, which is in turn a matter what a reasonable interpreter could take the common ground to be. If some common ground that does not licence the inference *that the speaker means c* could be taken by a reasonable interpreter to be the common ground of the conversation in which the relevant utterance is produced, then the speaker enjoys deniability over *c*. It seems then really easy for a speaker to enjoy deniability. Indeed, a speaker lacks deniability over her speaker-meant content *c* (she is *liable for* it) only if *every* plausible context is such that the speaker would be taken to mean *c*.[90] Camp further associates liability/deniability with *conversational record* in the following way. A speaker-meant content makes it into the conversational record if and only if the speaker is liable for it. By contrast, a speaker-meant content is 'off-record' just when the speaker enjoys deniability over it.

Camp does not think that the on-record or off-record status of a speaker-meant content aligns with primary and secondary speaker-meaning. One can lack deniability over a content one secondarily means (e.g. what one implies), and one can enjoy deniability over a content one primarily means (e.g. what one says). Some authors agree (Dinges & Zakkou 2023, García-Carpintero 2018, Peet 2015). Other authors disagree, associating liability with primariness and an absence of it with secondariness (Borg 2019, Fricker 2012). Now, if Camp

[89] Lee and Pinker associate deniability with indirectly conveyed contents, but they also think that there is a 'higher-order' form of deniability directed at a 'virtual audience' of 'outsiders and authorities' rather than at one's actual audience (Lee & Pinker 2010). The authors thus have in mind something similar to Camp when it comes to this latter form of deniability. One point of difference between them is that Lee and Pinker take higher-order deniability to be associated with the fact that the speaker's communicative intentions, even when recognised by the audience, do not become common ground. Camp argues that at least sometimes the speaker's communicative intentions do make it into the common ground while being deniable by the speaker.

[90] One may find in (Asher & Lascarides 2013) an alternative interpretation of the notion of *plausible context*. A context is plausible just when it makes the speaker's utterance *coherent* with previous discourse - the notion of coherence receiving theoretical treatment by the authors. A reasonable interpreter would then need to be someone attuned to discourse coherence. Whenever more than the consideration of coherence is needed to attribute an intended content to a speaker, the speaker enjoys deniability over her intended content.

is right, one cannot distinguish asserting from what one does in implying in terms of liability (on-record status). My present inclination is to agree with Camp. It seems to me that the liability issue has been mistakenly conflated with the issue of the commitment distinctive of assertion (Davies 2019). Conflating the two issues leads to major difficulties for some views of testimonial knowledge, as argued by Peet (2015). The package view that liability straddles the assertion-implication border while assertion comes with a distinctive commitment *à la* Viebahn seems to me consistent (Viebahn 2021).

I should further note that the notion of liability explored by Camp is not a purely theoretical construct. Camp shows that the standard she proposes roughly aligns with the standard operative in legal and legal-adjacent contexts. A speaker can be legally held responsible for having X'd content *c* (where X is some illocutionary force) just when there is no plausible context in which this person could be taken to have done something other than X-ing *c*. An illuminating illustration provided by Camp is the legal US standard for sexual harassment, and more specifically for *quid pro quo* sexual requests. The official standard is not whether the plaintiff took the suspect to request sexual favours with their utterances, but whether a 'reasonable person' could interpret the utterances otherwise than as requests for sexual favours (Camp 2018, pp.51-52).

This short summary of Camp's view does not reveal all its riches. But the crucial point for present purposes is that Camp's view of the conversational record is not consistent with the view I have sketched. I have suggested that what goes on the conversational record is determined by what a common-ground-defined audience would take the speaker to mean. Assuming Camp's connection between conversational record and liability, this seems to entail that a speaker is liable for a content *c* just when a common-ground-defined audience would take her to mean *c*. But the central aspect of Camp's view is precisely that the actual context in which the utterance is made - the actual common ground - is not relevant to what goes on the conversational record. Idealising somewhat, suppose that speaker and hearer make the same presuppositions. They are both right about the common ground, and this allows them to communicate successfully: the speaker X's that *p*, and the hearer takes the speaker to X that *p*. In such a case, the common-ground-defined audience is just the audience as it actually is. What Camp tells us is that even in such cases the speaker may enjoy deniability over *p*. My account of the conversational record does not allow this: that the speaker X'd that *p* goes on the conversational record just when a common-ground-defined audience would take the speaker to have X'd that *p*.

Rather than argue against Camp's account, I want to make it consistent with my view. To regain consistency, I must claim that we need two notions of conversational record: Camp's notion, which is closely related to what a speaker is legally (or legally-adjacent) liable for; and my notion, which is tied to ordinary reporting practices. There are then two conversational records: the *global* (or *extra-conversational*) record; and the *domestic* (or *intra-conversational*) record. When one reports speech acts in everyday situations, one tracks responsibilities determined by the actual context - common ground - of the speaker's utterance. The domestic record tracks these responsibilities. In legal and legal-adjacent

contexts one also tracks what speakers are responsible for, but the incurred responsibilities are determined by a range of non-actual contexts. The global record tracks these responsibilities. Both notions of conversational record are needed, since both our ordinary reporting practices and legal liability for one's speech stand in need of explanation.[91]

We saw in the previous chapter that the interpretation of a random virtuous interpreter is irrelevant to the quality of a speaker's communicative performance. Speakers craft their communicative performance for their audience as they think the common ground characterises it, rather than for a random virtuous interpreter.[92] Evaluating the speaker's communicative performance is then about evaluating the adequacy of such communicative tailoring. When it comes to deniability/liability, Camp's view effectively appeals to the interpretation of a random virtuous interpreter. However, her theory does not quite present itself as a simple modal condition on the interpretation of a virtuous interpreter. Her view is not that a speaker is liable for a content *c* just when a random virtuous interpreter would take the speaker to mean *c*. It is rather that a speaker is liable for *c* just when every context which a reasonable (virtuous) interpreter could take to be the context of the conversation leads this interpreter to take the speaker to mean *c*.

Couldn't a simple modal condition on the interpretation of a virtuous interpreter do the job of accounting for deniability/liability? First of all, it seems to me that the two conditions make different demands on virtuous interpreters. The simple modal condition does not demand that the virtuous interpreter attempt to build a plausible context in order to settle on an interpretation of the speaker's utterance. A random virtuous interpreter who is under no such constraint could simply refrain from attributing a definite communicative intention to the speaker, instead of venturing guesses about the common ground of the conversation in order to reach a definite interpretation. The same is true of my favoured modal condition and other modal conditions on the interpretation of an audience. Take *Venice Beach*. The verdict that the speaker's content-act is void originates in the fact that in most of the relevant possibilities, a

---

[91] Borg also introduces two notions of liability: *conversational* liability and *strict* liability (Borg 2019). Her notion of conversational liability aligns with my notion of domestic record to the extent that speakers are conversationally liable for the contents of pre-theoretical say-judgements. Her notion however diverges from mine to the extent that Borg does not take speakers to be conversationally liable for the content of secondary acts such as insinuations (or at least not to the same degree). As for Borg's notion of strict liability, it does not align with my notion of global record (Camp's notion of conversational record), since Borg takes strict liability to be uniquely associated with linguistically-determined content. This point is emphasised by Camp (2018, p.60). Simplifying somewhat, Borg's two notions of liability are designed to track two different kinds of content (what is asserted on the one hand, and what is linguistically determined on the other). My two notions of liability do not map onto specific types of content: both notions apply to various types of speech act content.

[92] Speakers who bear in mind the legal or public repercussions of their speech even when they address a particular audience in a non-legal setting might take into account the interpretation of a random virtuous interpreter. Such speakers engage in simultaneous double-crafting: communicative crafting for their core audience and legal crafting for random virtuous interpreters. They might devise their utterances so that on the one hand their core audience recognises their communicative intentions, and on the other hand they enjoy deniability over these communicative intentions. Political campaign speeches are often the locus of such double-crafting.

common-ground-defined audience refrains from attributing an intended content to the speaker's utterance given the paucity of the common ground and the vagueness of the speaker's gesture. Camp's proposed criterion for deniability/liability differs from all simple modal conditions on the interpretation of an agent to the extent that it demands that the agent reaches a definite interpretation.

It seems to me that this constraint goes hand in hand with the universal quantification in Camp's condition: once one forces reasonable interpreters to attribute a definite communicative intention to the speaker by making reasonable guesses about the context, one may demand that *every* reasonably-guessed context leads to the same interpretation. By contrast, the 'would' in a criterion for liability according to which *a speaker is liable for content c just when a random virtuous interpreter would take the speaker to mean c* had better not be interpreted as expressing universal quantification. As I have just noted, there are possibilities in the modal domain in which the virtuous interpreter refrains from attributing a definite communicative intention to the speaker. There are thus possibilities in which the virtuous interpreter does not reach a definite interpretation. According to a simple modal condition for liability, it is enough that there are such possibilities in the modal domain for a speaker not to be liable for a content *c*. This does not seem right. If 'would' expresses universal quantification, the simple modal condition overgenerates deniability.

One alternative is to have 'would' express that *most of* the relevant possibilities are ones in which a virtuous interpreter takes the speaker to mean *c* (see section 2 of the previous chapter). But first it seems to me unclear how one can establish the proportion of possibilities in which the interpreter refrains from attributing a definite communicative intention to the speaker. Independently of this problem, the 'most' reading of the simple modal condition would conceal which possibilities are relevant to deniability/liability. But we do know which possibilities are relevant. These are the possibilities in which the virtuous interpreter reaches a definite interpretation by constructing a plausible context for the utterance. Therefore, it seems to me that Camp's condition for deniability/liability cannot be made equivalent to a simple modal condition on the interpretation of a virtuous audience. However, Camp's condition can be turned into a virtuous audience modal condition, *with a further restriction on the modal domain*. According to this qualified virtuous audience condition, a speaker is liable for a content *c* just when a virtuous interpreter *who reaches a definite interpretation of the speaker's utterance* would take the speaker to mean *c*.

Conclusion

A modal condition on the interpretation of a common-ground-defined audience can serve to predict what a speaker says, despite some difficulties pertaining to the ordinary uses of 'say nothing' and 'say' more generally (section 1). This modal condition can also serve to predict what a speaker implies, as well as other speech act contents (section 2). By reporting the interpretation of a common-ground-defined audience, we track the content the speaker is responsible for, which is abstracted from the illocutionary responsibility incurred in producing

the utterance. Tracking the responsibilities incurred by successive utterances in a conversation is not tracking the *conversational score* in any acceptable sense (section 3). These views are consistent with several theories of illocutionary force, and I regard them as leading to a pluralist stance on the *conversational record* (section 4).

Bibliography

Åkerman, J. (2009). ‘A plea for pragmatics’. *Synthese* 170: 155–167.

Åkerman, J. (2015). ‘The communication desideratum and theories of indexical reference’. *Mind and Language* 30 (4):474–499.

Asher, N. and Lascarides, A. (2013). ‘Strategic Conversation’. *Semantics & Pragmatics* 6: 1-62.

Austin, J.L. (1957). ‘A plea for excuses’. *Proceedings of the Aristotelian Society*. 57:1-30.

Austin, J.L. (1975). *How to do things with words*. Second edition. Edited by Marina Sbisá & J. O. Urmson. Oxford: Clarendon Press.

Bach, K. & Harnish, R. (1979). *Linguistic Communication and Speech Acts*. MIT Press.

Bach, K. (1992). ‘Intentions and Demonstrations’. *Analysis* 52 (3):140-146.

Bach, K. (1994a). *Thought and Reference*. Second edition. Oxford University Press.

Bach, K. (1994b). ‘Conversational Impliciture’. *Mind and Language* 9(2), pp. 124–162.

Bach, K. (2001). ‘You don’t say?’. *Synthese*. 128: 15–44.

Bach, K. (2005). 'Context ex Machina'. In Zoltan Szabo (ed.) *Semantics v. Pragmatics*, pp.15-44. Oxford University Press.

Bach, K. (2017). ‘Reference, Intention, and Context: Do Demonstratives Really Refer?’ in M. de Ponte and K. Korta (eds) *Reference and Representation in Thought and Language*, pp.57–72. Oxford University Press.

Baier, A.C. (1970). ‘Act and intent’. *Journal of Philosophy* 67:19, 648-658.

Baron-Cohen, S. (1995). *Mindblindness. An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press

Bertolet, R. (1987). ‘Speaker Reference’. *Philosophical Studies* 52:2, 199-226.

Bloom, P. (2001). ‘Précis of how children learn the meanings of words’. *Behavioral and Brain Sciences* 24 (6):1095-1103.

Borg, E. (2004). *Minimal Semantics*. Oxford University Press.

Borg, E. (2019). ‘Explanatory roles for minimal content’. *Noûs* 53(3), 513–539.

Brandom, R. (1983). ‘Assertion’. *Noûs* 17(4), 637–650

Brandom, R. (2008). *Between saying and doing: towards an analytic pragmatism*. New York: Oxford University Press.

Bratman, M. (1984). ‘Two Faces of Intention’. *The Philosophical Review.* 93:3, 375-405.

Bratman, M. (1990). *Intentions, Plans and Practical Reasons*. CSLI Publications, Stanford.

Braun, D. (1996). ‘Demonstratives and their linguistic meanings’. *Noûs* 30 (2):145-173.

Braun, D. (2011). ‘Implicating Questions’. *Mind and Language* 26 (5):574-595.

Byrne RW, Cartmill E, Genty E, Graham KE, Hobaiter C, Tanner J. (2019). ‘Great ape gestures: intentional communication with a rich set of innate signals’. *Animal Cognition* 20:755–769.

Camp, E. (2006): ‘Contextualism, Metaphor, and What is Said’. *Mind and Language* 21:3, 280-309.

Camp, E. (2007): ‘Prudent Semantics Meets Wanton Speech Act Pluralism’. In G. Preyer and G. Peter (eds.) *Context-Sensitivity and Semantic Minimalism: New Essays on Semantics and Pragmatics*. pp.194-213. Oxford: Oxford University Press.
Camp, E. (2011). ‘Sarcasm, Pretense, and The Semantics/Pragmatics Distinction’. *Noûs* 46 (4):587-634.
Camp, E. (2018). ‘Insinuation, common ground, and the conversational record’. In D. Fogal, D. Harris & M. Moss (eds.) *New work on speech acts*. pp.40–67. Oxford: Oxford University Press.
Campbell, L. (2019). ‘Propositionalism about intention: shifting the burden of proof’. *Canadian Journal of Philosophy* 49:2, 230-252.
Cappelen, H. and Lepore, E. (1997). ‘On an Alleged Connection Between Indirect Speech and the Theory of Meaning’. *Mind and Language* 12:3/4, 278–296.
Cappelen, H. and Lepore, E. (2005). *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Malden, MA: Wiley-Blackwell.
Carruthers, P. (2006). *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought*. New York: Oxford University Press UK.
Carston, R. (2021). ‘Polysemy: Pragmatics and sense conventions’. *Mind and Language* 36 (1):108-133.
Carston, R. (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.
Castañeda, H.N. (1971). ‘Intentions and the Structure of Intending’. *The Journal of Philosophy* 68:15, 453- 466.
Chisholm R.M. (1970). ‘The Structure of Intention’. *The Journal of Philosophy*. 67:19 633-647.
Coltheart, M. (1999). ‘Modularity and Cognition’. *Trends in Cognitive Sciences* 3:3 115-120.
Dancy, J. (2017). ‘Moral Particularism’. In E.N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition).
Davidson, D. (1967). ‘Truth and meaning’. *Synthese* 17 (1):304-323.
Davidson, D. (1986). ‘A nice derangement of epitaphs’. In R. Grandy and R. Warner (eds), *Philosophical Grounds of Rationality: Intentions, Categories, Ends*. Oxford: Clarendon Press.
Davies, A. (2019). ‘Testimony, recovery and plausible deniability: A response to Peet’. *Episteme* 16 (1):18-38.
Davis, W.A. (1998). *Implicature: Intention, Convention, and Principle in the Failure of Gricean Theory*. New York: Cambridge University Press.
De Villiers, J. Stainton, R. J. & Szatmari, P. (2007). ‘Pragmatic abilities in autism spectrum disorder: A case study in philosophy and the empirical’. *Midwest Studies in Philosophy* 31 (1):292–317.
Devitt, M. (2022). ‘The irrelevance of intentions to refer: demonstratives and demonstrations’. *Philosophical Studies* 179 (3):995-1004.
Dinges, A. & Zakkou, J. (2023). ‘On Deniability’. *Mind* 132 (526):372-401.
Elbourne, P. (2008). ‘Demonstratives as individual concepts’. *Linguistics and Philosophy* 31 (4):409-466.
Elbourne, P. (2020). ‘Weather predicates, binding, and radical contextualism’. *Mind and Language* 37 (1):56-72.

Evans, G. (1982). *Varieties of Reference*. Oxford University Press.
Ferrero, L. (2013). 'Can I Only Intend My Own Actions?'. In *Oxford Studies in Agency and Responsibility*. Oxford: Oxford University Press.
Fitch W.T. (2010). *The Evolution of Language*. Cambridge: Cambridge University Press.
Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
Fricker, E. (2012). 'Stating and insinuating'. *Proceedings of the Aristotelian Society* 86, 61–94.
Frith U. and Happé F. (1994), 'Language and Communication in Autistic Disorders'. *Philosophical Transactions of the Royal Society of London* 346: 97–104.
García-Carpintero, M. (1998). 'Indexicals as token-reflexives'. *Mind* 107 (427):529-564.
García-Carpintero, M. (2018). 'Sneaky Assertions'. *Philosophical Perspectives* 32(1): 188-218.
García-Carpintero, M. (forthcoming). 'Intention and Convention Revisited'. *Analytic Philosophy*.
Gauker, C. (2008). 'Zero tolerance for pragmatics'. *Synthese* 165 (3):359–371.
Gauker, C. (2018). 'Against the speaker-intention theory of demonstratives'. *Linguistics and Philosophy* 42 (2): 109-129.
Gazdar, G. (1981). 'Speech act assignment'. In A. Joshi, B. L. Webber & I. Sag (eds.), *Elements of discourse understanding*, pp. 64–83. Cambridge: Cambridge University Press.
Geurts B. & Rubio-Fernandez P. (2015). 'Pragmatics and Processing'. *Ratio* 28 (4):446-469.
Geurts, B. (2019). 'Communication as Commitment Making: Speech Acts, Implicatures, Common Ground'. *Theoretical Linguistics* 45 (1–2): 1–30.
Gillette, J., Gleitman, H., Gleitman, L. & Lederer, A. (1999). 'Human simulations of vocabulary learning'. *Cognition* 73 135-176.
Glanzberg, M. (2007). 'Context, Content and Relativism'. Philosophical Studies 136 (1): 1-29.
Goldberg, S.C. (2015). *Assertion: On the Philosophical Significance of Assertoric Speech*. New York, NY: Oxford University Press.
Goldberg, S.C. (2023). 'Illocutionary Force, Speech Act Norms, and the Coordination and Mutuality of Conversational Expectations'. In L. Caponetto & P. Labinaz (eds.). *Sbisà on Speech as Action*. pp.195-219. Palgrave-Macmillan.
Gomez-Torrente, M. (2019). *Roads to Reference: An Essay on Reference Fixing in Natural Language*. Oxford University Press.
Green, M. (2013). 'Assertions'. In M. Sbisà and K. Turner (eds.). *Pragmatics of Speech Actions*. pp.387–409. Berlin: De Gruyter Mouton.
Green, M. (2021). 'Speech Acts'. In E.N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy*. Fall 2021 Edition.
Grice, H.P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
Hamblin, C.L., 1973, 'Questions in Montague English'. *Foundations of Language* 10 (1): 41–53.
Happe, F., Brownell, H. and Winner, E. (1999). 'Acquired 'theory of mind' impairments following stroke'. *Cognition* 70 211–40.
Harman, G. (1977). 'How to Use Propositions'. *American Philosophical Quarterly* 14: 173–176.

Harris, D. W. (2020). ‘Semantics without semantic content’. *Mind and Language* 37 (3):304-328.
Heck, R.K. (2014). ‘Semantics and Context-Dependence: Towards a Strawsonian Account’. In B. Sherman & A. Burgess (eds.). *Metasemantics: New Essays on the Foundations of Meaning*. pp. 327-364. Oxford University Press.
Heim, I. & Kratzer, A. (1998). *Semantics in generative grammar*. Malden, MA: Blackwell.
Jacobson, P. (1999). ‘Towards a variable-free semantics’. *Linguistics and Philosophy* 22 (2):117-185.
Jeshion, R. (2010). ‘Singular Thought: Acquaintance, Semantic Instrumentalism, and Cognitivism’. In Robin Jeshion (ed.) *New Essays on Singular Thought.* pp. 105-141. Oxford University Press.
Kaplan, D. (1978). ‘Dthat’. In Peter Cole (ed.). *Syntax and Semantics*. pp. 221-243. Academic Press.
Kaplan, D. (1989). ‘Demonstratives’. In J. Almog, J. Perry & H. Wettstein (eds.) *Themes From Kaplan*. pp.481-563. Oxford University Press.
Kaplan, D. (1989b). ‘Afterthoughts’. In J. Almog, J. Perry & H. Wettstein (eds.) *Themes From Kaplan*. pp. 565-614. Oxford University Press.
King, J.C. & Stanley, J. (2005). ‘Semantics, pragmatics, and the role of semantic content’. In Z.G. Szabo (ed.). *Semantics Versus Pragmatics*. pp.111-164. Oxford University Press.
King, J. (2013). ‘Supplementives, the coordination account, and conflicting intentions’. *Philosophical Perspectives* 27(1): 288–311.
King, J. (2014). ‘Speaker intentions in context’. *Noûs* 48(2) 219–237.
King, J. C. (2001). *Complex Demonstratives: A Quantificational Account*. Cambridge, MA, USA: MIT Press.
Kissine, M. (2012). ‘Pragmatics, Cognitive Flexibility and Autism Spectrum Disorders’. *Mind and Language* 27 (1):1-28.
Kölbel, M. (2011). ‘Conversational Score, Assertion, and Testimony’. In Jessica Brown & Herman Cappelen (eds.), *Assertion: New Philosophical Essays*. pp. 49-77. Oxford University Press.
Korta, K. & Perry, J. (2011). *Critical Pragmatics: An Inquiry Into Reference and Communication*. New York: Cambridge University Press.
Kripke, S. (1977). ‘Speaker’s Reference and Semantic Reference’. *Midwest Studies in Philosophy* 2 (1):255-276.
Kukla, Q. & Lance, M. (2009). *‘Yo!’ and ‘Lo!’: The Pragmatic Topography of the Space of Reasons*. Harvard University Press.
Langdon, R.; Davies, M. & Coltheart, M. (2002). ‘Understanding Minds and Understanding Communicated Meanings in Schizophrenia’. *Mind and Language* 17 (1‐2):68-104.
Larson, R. & Segal, G. (1995). *Knowledge of Meaning: An Introduction to Semantic Theory*. MIT Press.
Lawlor, K. (2005). ‘Confused thought and modes of presentation’. *Philosophical Quarterly.* 55 (218): 21-36.
Lawlor, K. (2007). ‘A notional worlds approach to confusion’. *Mind & Language* 22 (2): 150–172.

Lewerentz, L. & Marschall, B. (2018). ‘Metasemantics, intentions and circularity’. *Synthese* 195(1): 1667–1679.
Lewis, D.K. (1970). ‘General semantics’. *Synthese* 22 (1-2):18-67.
Lewis, D. (1979). ‘Scorekeeping in a language game’. *Journal of Philosophical Logic* 8 (1):339-359.
Lewis, D. (1980). ‘Index, context, and content’. In S. Kanger & S. Ohman (eds.). *Philosophy and Grammar*. pp. 79– 100. Amsterdam: Reidel.
MacFarlane, J. (2011). ‘What is assertion?’. In H. Cappelen & J. Brown (eds.). *Assertion: New philosophical essays*. pp.79–96. Oxford: Oxford University Press.
Marsili, N. & Löhr, G. (2022). ‘Saying, commitment, and the lying-misleading distinction’. *Journal of Philosophy* 119 (12):687-698.
Marsili, N. (2023). ‘Towards a Unified Theory of Illocutionary Normativity’. In L. Caponetto & P. Labinaz (eds.) *Sbisà on Speech as Action*. pp.165-193. Palgrave-Macmillan.
Marsili, N. (forthcoming). ‘The Definition of Assertion: Commitment and Truth’. *Mind and Language*.
McGinn, C. (1981). 'The Mechanism of Reference'. *Synthese* 49 157-86.
Mele, A. R. (1992). *Springs of action*. Oxford University Press.
Millikan, R. G. (2000). *On Clear and Confused Ideas: An Essay about Substance Concepts*. Cambridge University Press.
Montminy, M. (2010). ‘Context and Communication: A Defense of Intentionalism’. *Journal of Pragmatics* 42, 2910-2918.
Moore, R. (2017). ‘Gricean Communication and Cognitive Development’. *Philosophical Quarterly* 67 (267):303-326.
Mount, A. (2008). ‘Intentions, gestures, and salience in ordinary and deferred demonstrative reference’. *Mind and Language* 23 (2):145–164.
Nagano M., Zane E., Grossman R.B. (2021). ‘Structural and Contextual Cues in Third-person Pronoun Interpretation by Children with Autism Spectrum Disorder and their Neurotypical Peers’. *Journal of Autism and Developmental Disorders* 51(5): 1562–1583.
Neale, S. & Schiffer, S. (2020). ‘How Demonstratives and Indexicals Really Work’. In S. Biggs and H.Geirsson (eds). *The Routledge Handbook of Linguistic Reference*. Routledge.
Neale, S. (1992). ‘Paul Grice and the Philosophy of Language’. *Linguistics and Philosophy*. 15 (5), 509-559.
Neale, S. (2004). ‘This, That and The Other’. In M. Reimer and A. Bezuidenhout (eds.). *Descriptions and Beyond*. pp.68–182. Oxford: Clarendon Press.
Neale, S.(2016). ‘Silent Reference’, in Gary Ostertag (ed.). *Meanings and Other Things: Themes From the Work of Stephen Schiffer*. Oxford University Press.
Norbury, C.F. (2005). ‘Barking up the wrong tree? Lexical ambiguity resolution in children with language impairments and autistic spectrum disorders’. *Journal of Experimental Child Psychology* 90 (2), 142-171.
Novogrodsky R, & Edelson L.R. (2016). ‘Ambiguous pronoun use in narratives of children with Autism Spectrum Disorders’. *Child Language Teaching and Therapy* 32(2) 241–252.
Nowak, E. & Michaelson, E. (2020). ‘Discourse and method’. *Linguistics and Philosophy* 43 (2):119-138.

Nowak, E. & Michaelson, E. (2021). ‘Who’s Your Ideal Listener?’. *Australasian Journal of Philosophy* 99(2): 257-270.
Nowak, E. & Michaelson, E. (2022). ‘Meta-Metasemantics, or the Quest for the One True Metasemantics’. *The Philosophical Quarterly* 72 (1): 135–154.
Nowak, E. (2019). ‘No context, no content, no problem’. *Mind and Language* 36 (2):189-220.
Pacherie, E. (2008), ‘The phenomenology of action: a conceptual framework’. *Cognition* 107(1), 179–217.
Pagin, P. & Westerståhl, D. (2010a).‘Compositionality I: Definitions and Variants’, *Philosophy Compass* 5(3): 250–264.
Pagin, P. & Westerståhl, D. (2010b). ‘Compositionality II: Arguments and Problems’. *Philosophy Compass* 5 (3):265-282.
Peet, A. (2015). ‘Testimony, pragmatics, and plausible deniability’. *Episteme* 12(1), 29–51.
Peet, A. (2021). ‘Assertoric content, responsibility, and metasemantics’. *Mind and Language* 37 (5):914-932.
Perry, J. (1997). ‘Indexicals and Demonstratives’. In Bob Hale & Crispin Wright (eds.) *A Companion to the Philosophy of Language*. pp.486-612. Blackwell.
Perry, J. (2009). ‘Directing intentions’. In J. Almog & P. Leonardi (eds.) *The Philosophy of David Kaplan*. pp.187-201. Oxford University Press.
Predelli, S. (2010). ‘Malapropisms and the simple picture of communication’. *Mind and Language* 25 (3):329-345.
Rabern, B. (2012). ‘Against the identification of assertoric content with compositional value’. *Synthese* 189 (1):75-96.
Radulescu, A. (2019). ‘A Defence of Intentionalism about Demonstratives’. *Australasian Journal of Philosophy* 97 (4): 775-791.
Recanati, F. (1986). ‘On Defining Communicative Intentions’. *Mind and Language* 1 (3):213-41.
Recanati, F. (2001). ‘Déstabiliser le sens’. *Revue Internationale de Philosophie* 2 (217):197-208.
Recanati, F. (2004). *Literal Meaning*. Cambridge University Press.
Recanati, F. (2010). *Truth-conditional Pragmatics*. Oxford University Press.
Recanati, F. (2012). *Mental Files*. Oxford University Press.
Recanati, F. (2016). *Mental Files in Flux*. Oxford University Press.
Reiland, I. (2023). ‘Linguistic Mistakes’. *Erkenntnis* 88 (5):2191-2206.
Reimer, M. (1991). ‘Demonstratives, demonstrations, and demonstrata’. *Philosophical Studies* 63, 187–202.
Reimer, M. (1992). ‘Three views of demonstrative reference’. *Synthese*. 93(3): 373–402.
Reimer, M. (2004). ‘What malapropisms mean: A reply to Donald Davidson’. *Erkenntnis* 60 (3):317-334.
Robbins, P. (2007). ‘Minimalism and Modularity’. In G. Preyer & G. Peter (eds.) *Context-Sensitivity and Semantic Minimalism: New Essays on Semantics and Pragmatics*. Oxford University Press.
Robbins, P. (2017). ‘Modularity of Mind’. In E.N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition).

Roberts, C. (2002). ‘Demonstratives as Definites’, In K. van Deemter & R. Kibble (eds.). *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*. pp. 89-196. CSLI Press..
Roberts, C. (2018). 'Speech Acts in Discourse Context', in D. Fogal, D.W. Harris, and M.Moss (eds). *New Work on Speech Acts.* pp. 317–359. Oxford: Oxford University Press.
Rostworowski, W. ; Kuś, K. & Maćkiewicz, B. (2022). ‘Against intentionalism: an experimental study on demonstrative reference’. *Linguistics and Philosophy* 45 (5):1027-1061.
Saba, W. S. & Corriveau, J.P. (2001). ‘Plausible reasoning and the resolution of quantifier scope ambiguities’. *Studia Logica* 67 (2):271-289.
Sainsbury, M. (2020). 'Varieties of Singularity'. In R. Goodman, J. Genone, and N. Kroll (eds). *Singular Thought and Mental Files*. pp.21-37. Oxford University Press.
Saul, J.M. (2001). ‘Wayne A. Davis, Implicature: Intention, convention, and principle in the failure of Gricean theory’. *Noûs* 35 (4):631-641.
Saul, J.M. (2002). ‘Speaker meaning, what is said, and what is implicated’. *Noûs* 36 (2):228–248.
Saul, J.M. (2010). ‘Speaker-meaning, Conversational Implicature and Calculability’. In K. Petrus (ed.) *Meaning and analysis: new essays on Grice*. New York: Palgrave-Macmillan.
Sbisà, M. (1984). ‘On Illocutionary Types’. *Journal of Pragmatics* 8: 93-112.
Schiffer, S. (1972). *Meaning*, Oxford University Press.
Schoubye, A.J. (2011). ‘Ghosts, Murderers, and the Semantics of Descriptions’. *Noûs* 47 (3):496-533.
Searle, J.R. (1968). ‘Austin on locutionary and illocutionary acts’. *Philosophical Review* 77 (4):405-424.
Searle, J.R. (1969). *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
Searle, J. (1980). ‘The Intentionality of Intention and Action’. *Cognitive Science* 4, 47-70.
Shapiro, L. (2020). ‘Commitment accounts of assertion’. In S. Goldberg (ed.). *The Oxford Handbook of Assertion*. pp.73–97. Oxford: Oxford University Press.
Soames, S. (2009). *Philosophical Essays*.Volume 1, Princeton University Press.
Speaks, J. (2016). ‘The Role of Speaker and Hearer in the Character of Demonstratives’. *Mind*. 125 (498): 301-339.
Speaks, J. (2017). ‘A puzzle about demonstratives and semantic competence’. *Philosophical Studies* 174 (3):709-734.
Sperber, D. & Wilson, D. (1995). *Relevance: Communication and Cognition*. Oxford: Blackwell (2nd edition).
Sperber, D. & Wilson, D. (2002). ‘Pragmatics, Modularity and Mind‑reading’. *Mind and Language* 17 (1-2):3–23.
Stainton, R. (2016). ‘Full-on stating’. *Mind & Language* 31(4): 395–413.
Stalnaker, R. (1999). ‘Assertion’, reprinted in *Context and Content*. pp.78–95. Oxford University Press,
Stalnaker, R. (2002). ‘Common Ground’. Linguistics and Philosophy, 25 (5-6), 701-721.
Stalnaker R. (2014). *Context*. Oxford University Press.

Stanley, J. (2005). ‘Review of François Recanati, Literal Meaning’. *Notre Dame Philosophical Reviews* (9).
Stanley, J. (2007). *Language in context: selected essays*. New York: Oxford University Press.
Stojnic, U., Stone, M., & Lepore, E. (2013). ‘Deixis (even without pointing)’. *Philosophical Perspectives* 27(1), 502–525.
Stojnic, U., Stone, M., & Lepore, E. (2017). ‘Discourse and logical form: Pronouns, attention, and coherence’. *Linguistics and Philosophy* 40(5), 519–547.
Stokke, A. (2010). ‘Intention-sensitive semantics’. *Synthese* 175(3): 383–404.
Strawson, P. F. (2000). ‘Direct Singular Reference: Intended reference and Actual Reference’. In *Entity and Identity and Other Essays*. pp.92-99. Oxford: Clarendon Press.
Swaab, T., Brown, C., Hagoort, P. (2003). ‘Understanding words in sentence contexts: the time course of ambiguity resolution’. *Brain and Language* 86 (2):326–343.
Szabó, Z. G. (2020). ‘Compositionality’. In E.N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition).
Tager-Flusberg, H., Sullivan K. (2000), ‘A Componential Theory of Mind: Evidence from Williams Syndrome’. *Cognition* 76: 59–89.
Thompson, R.J.(2014). ‘Meaning and Mindreading’. *Mind and Language* 29 (2):167-200.
Tomasello, M. (2010). ‘Perceiving intentions and learning words in the second year of life’. In Bowerman, M., & Levinson, S. (eds.). *Language Acquisition and Conceptual Development*. pp.132-158. Cambridge University Press.
Tomasello, M. and Call, J. (2019). ‘Thirty years of great ape gestures’. *Animal Cognition* 22:461–469.
Unnsteinsson, E. (2016). ‘Confusion is Corruptive Belief in False Identity’. *Canadian Journal of Philosophy* 46 (2):204-227.
Unnsteinsson, E. (2017). ‘A Gricean Theory of Malaprops’. *Mind and Language* 32 (4):446-462.
Unnsteinsson, E. (2019). ‘The Edenic Theory of Reference’. *Inquiry*, 62 (3):276-308.
Viebahn, E. (2017). ‘Non-literal Lies’. *Erkenntnis* 82 (6):1367-1380.
Viebahn, E. (2020). ‘Ways of Using Words: On Semantic Intentions’. *Philosophy and Phenomenological Research* 100 (1):93-117.
Viebahn, E. (2021). ‘The lying–misleading distinction: A commitment based approach’. *The Journal of Philosophy* 118 (6):289-319.
Wettstein, H. (1984). ‘How to bridge the gap between meaning and reference’. *Synthese* 58 (1):63-84.
Zempleni, M. et al. (2017). ‘Semantic ambiguity processing in sentence context: Evidence from event-related fMRI’. *NeuroImage* 34 (3): 1270-1279.
Ziff, P. (1972): ‘What Is Said’. In D. Davidson and G. Harman (eds.). *Semantics of Natural Language*. pp. 709–721. Dordrecht: Reidel.