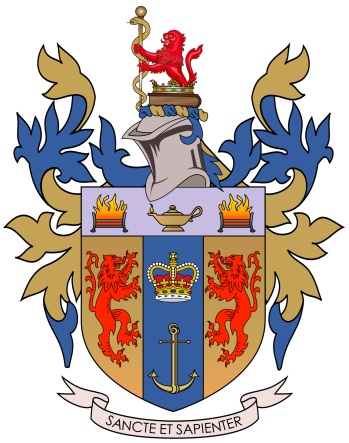
# On the unity of intentional content

PhD thesis in Philosophy

Hugo Heagren

King's College London



This page intentionally not blank.

To four women without whom it would not have been possible: Sandy, Ruth, Giulia and Colette.

# Acknowledgements

Thanks first to my supervisors, Mark Textor and Eliot Michaelson. After four years, I am still regularly astonished at the breadth of Mark's reading and his capacity to understand, immediately, the parts of my writing which I have least understood myself. I am continually (pleasantly!) surprised by what controversies Eliot takes to be obvious, and what 'obvious' claims he thinks require more. Together you have somehow both kept me on my toes and brought me down to earth.

My most heartfelt thanks go to those in the philosophical community who have been there along the way. To Grace Benita, Bill Brewer, Susie Clark, Niklas Dahl, Dan Dearlove, Josh Dever, Giulia Felappi, Alex Geddes, Alex Grzankowski, Ed Hornby-King, Will McNeil, Jessie Munton, Collete Olive, Alex Oliver, Juuso Rantanen, Josie Roux, Mark Sainsbury, Matt Soteriou, James Stazicker, Lee Walters, and all those in the graduate communities of King's College London and UT Austin.

Thanks also, of various kinds, to those without whose care, attention and quiet belief I am sure I would not have found my way to philosophy at all. To Liz and John, and David and Wendy (all from around Sedbergh) and to my teachers: Mrs Cullen, Mrs Oughton and Chris Chambers (both of Sedbergh), Mr Clifton, Mr Middleton, Mr Parsons, Mr Sanger, Mr Sharp, Mrs Coley-Masters, Ms Jones, Ms Tromans and Mr Turner (all of King Edward VI Five Ways at one time). I am sure I was a terrible student of what much you could have taught me, but I have found my way here in the end.

And finally, to everyone not already mentioned: Harri Compton, Anton Friese, Siân Harris, Sionna Hurley-O'Kelly, Oliver Keyland, Finn Kinsler-O'Sullivan, Alex O'Leary, Jordan Roche and surely many others. Blue skies to everyone at Haddenham, and across the rest of FSC, who reminded me that there is a world outside academia from time to time. And of course to my family, especially Morgan and Isobel. Thank you.

# Abstract

Many philosophers have presumed something like:

**Unity** *The nature of the significance of names, perceptual experiences and singular thoughts is the same. This shared significance is often called 'content'.*

This presumption has gone largely undefended, though it has been relied on in a variety of philosophical work. The purpose of this dissertation is to provide good reasons to think it is true, and to draw out its immediate and significant consequences. This latter goal requires three separate theories: of what has content, of what content is, and of how it is determined what content is had by which things which have it. I devote a chapter to each.

Chapter 1 introduces the terms of the project, and situates it with respect to some recent figures.

Chapter 2 offers a good reason to believe Unity and also functions as a historical survey. Summarising the literature on these issues since Frege, I divide the examples and cases which arise in these discussions into four groups, and show that each kind of case arises in discussions of all three phenomena. On this basis I offer a methodological argument for Unity. I believe this is the first such argument.

In Chapter 3 I argue for a view of what things *have* content. This is important because *prima facie*, names, thoughts and perceptual experiences are very different kinds of thing, and it would be surprising if they bore identical relations to the world. In fact I show that names, thoughts and experiences are not the things which really bear content and that the things which do are not so different after all: utterances of names, cognitive tokens of concepts, and 'object files'.

Chapter 4 describes my theory of the nature of the 'significance' in Unity: something's significance is a combination of its referent (if any), and the properties an agent judges to hold of that referent. I show that when properly understood, the account handles well all the existing concerns across the literature.

This kind of account is not new. Those who have held similar views (and even those who have argued against them) have all assumed that on such a view, *what* referent something has is determined by the satisfaction of the properties

believed to hold of that referent. This principle has been the basis of many arguments against views like mine and I reject it. But this leaves a gap in my account: if reference is not determined by satisfaction, how is it determined? In Chapter 5 I close with a discussion of what fixes each aspect of the significance I discussed in the previous chapter. My theory is an elaboration of the ideas of 'causal' theorists, including Kripke, Evans and Sainsbury.

# Contents

# Chapter 1

# Introduction

Many philosophers have presumed something like:

**Unity** *The nature of the significance of names, perceptual experiences and singular thoughts is the same.*

Some examples. Russell (1905) says he will offer a theory of denoting phrases. Until the penultimate paragraph, the paper is concerned very strictly with only this. There Russell writes:

> One interesting result of the above theory of denoting is this: when there is anything which we do not have immediate *acquaintance*, but only definition by denoting phrases, then the propositions in which this thing is introduced by means of a denoting phrase do not really contain this thing as a constituent, but contain instead the constituents expressed by the several words of the denoting phrase. Thus in every proposition that we can *apprehend* (i.e. not only in those whose truth or falsehood we can *judge of*, but in all that we can *think about*), all the constitutuents are really entities which which we have immediate *acquaintance*.
>
> (Russell 1905 p. 492)

Denoting phrases and the words which constitute them are linguistic entities ( underlined). Acquaintance and thinking about are mental relations (*italics*). In this passage Russell moves freely between talk of language and cognition. His eventual claim assumes that the structure of the phrases (really the sentences *containing* such phrases) we use to express our thoughts is indicative of the relational nature of the thoughts themselves; Russell assumes that language and thought relate to their objects in the same way. McDowell (1982 p. 304) relies heavily on these comments.

Martin (2002a p. 2) ascribes a 'conception of thought and representation' to Gareth Evans according to which: if 'an intentional content is directed on a particular object or event and is 'singular" then 'such a content is 'object-dependent': one can think thoughts of this kind only where an appropriate object to be thought about actually exists'. This is clearly a view about *thoughts*. Martin goes on immediately:

> Since hallucinations are taken to be paradigm examples of the absence of any such candidate object of reference, object-dependent accounts of intentional content will ascribe no intentional content to hallucinations.

Hallucinations are perceptual phenomena (in the broad sense of the term—they have to do with the perceptual system). Martin concludes that the view attributed to Evans and McDowell (about thoughts) is committed to a view about hallucinations (about perception). So he assumes that holding a view about the nature of the content of thoughts commits one to a similar view about the content of perception.

This is even more explicit on p. 3. Martin says he will 'outline the doctrine of object dependence for singular thought and isolate the key element which presents the problem for intentional theories of perception'—so he thinks that a doctrine about the contents of certain *thoughts* presents a problem for certain theories of *perception*.

Schellenberg (2016) discusses a view she calls 'generalism' which claims 'perceptual states are constituted only by general elements'. She says:

> Generalism is attractive for a number of reasons [but] it comes at a cost. It does not give an adequate explanation of how it is that we single out particulars in perception by discriminating them from other particulars, thereby yielding perceptual knowledge of particulars, justifying singular thoughts, grounding demonstrative reference, fixing the reference of singular terms, and providing the basis for singular thoughts.
>
> (Schellenberg 2016 p. 27)

The first expected yield is uncontroversial—if anything provides perceptual knowledge of particulars, surely it is perception. But the rest are not. Schellenberg doesn't say why we should expect perceptual states to justify or provide the basis for singular thoughts, ground demonstrative reference, of fix the reference of singular terms. She just assumes that the role perception plays with respect to

perceptual knowledge is the same as that which it plays with respect to singular thought, and linguistic semantic content; i.e. something like Unity.

These are only three examples—there are many more.[1] The presumption of Unity has gone largely undefended, though it has been relied on in a variety of philosophical work. The purpose of this dissertation is to provide good reasons to think it is true, and to draw out its immediate and significant consequences. This latter goal requires three separate theories: of what has content, of what content is, and of how it is determined what content is had by which things which have it. I devote a chapter to each.

In the rest of this introduction I introduce the broad terms of the debate, describe briefly the shape of the arguments and theories, and explain what is innovative about them in the context of what has gone before. I close with some clarificatory notes on the scope of the project.

## 1.1 On contents

Some things have a certain kind of significance. Among those which do not are pianos, birds, bath towels and bookcases. Among those which plausibly do are names, pronouns, demonstratives, gestures of pointing, thoughts, perceptual experiences, and the like. There are also less clear candidates, like thermostats and portraits—some would put these on one side, some on the other and some would say they are borderline cases.

Different types of thing, and different tokens of each type, might have different kinds of significance. Tokens of various linguistic types (especially singular terms) refer or denote or designate; some connote or ascribe; some have been said to have 'semantic value', 'cognitive value', 'sense', 'ingredient sense' or 'semantic power'. Thoughts have also been said to refer, to be 'about' things, to have (intentional) objects and to have 'cognitive value'. Objective perceptual experiences have *also* been said to refer, to have (intentional) objects, 'particularity' or 'semantic features' and to 'convey' something to their subject. Finally, it is striking that singular terms, thoughts and perceptual experiences have all been said to be 'of' things.

Variations on the term 'content' have been used for all such kinds of significance in all three phenomena. These uses can be divided roughly into two groups. The first I will call *external* content: this is the significance a thing

[1] Evans (1973 pp. 197–8), Chisholm (1977 p. 68), Kim (1977), Evans (1982 p. 66), Zalta (1988 p. 110), Smith and Bittner (2001 p. 9), Zalta (2001 p. 343), Johnston (2004 pp. 148–50), S. Siegel (2005 §3.1), Bach (2010 p. 59), Crane (2011b 22ff; 2013 p. 139), Sainsbury and Tye (2012 Chapter 1), Beck (2014 pp. 563–6), Gauker (2015 p. 36), Goodman (2016), Jackman (2020 p. 27), and Shahmoradi (2024).

has by its connection with a worldly object (or for some theorists, its merely *apparent* connection). It is widely agreed that names, some descriptions and demonstratives (or utterances of these) refer to, denote or designate objects. For example, Devitt (1981) says 'Designation usually holds between proper names and the objects they refer to; it often holds between demonstratives, pronouns, and definite descriptions, and the objects they refer to'. Similarly Michaelson and Reimer (2022) write:

> Reference is a relation that obtains *between a variety of representational tokens and objects or properties*. For instance, when I assert that "Barack Obama is a Democrat," I use a particular sort of representational token—i.e. the name 'Barack Obama'—which refers to a particular individual—i.e. Barack Obama.
>
> (my emphasis)

Here the representational token which *has* the content is the name 'Barack Obama' (or the utterance or inscription of that name), and this token relates to a worldly object—Barack Obama himself. It is also widely agreed that some thoughts relate to objects in a similar way. Michaelson and Reimer continue:

> names and other referential terms are hardly the only type of representational token capable of referring (consider, for instance, concepts, mental maps, and pictures)

Crane (2013 p. 3) says similarly 'whenever someone thinks, there is something they think about' and continues immediately 'Clearly some perceptual experiences also relate us to the objects of the world—we can detect whole, persisting objects by (at least) seeing or touching them'. There are many other examples. So linguistic, mental and perceptual tokens can all have similar kinds of relation to worldly *things*. I call this 'external' content because the objects that language, thought and perception relate to are external to us; they are in the world 'out there'.

Contrast with what I will call *internal* content. This is also widely attributed to names, thoughts and perceptual experiences. It is often characterised negatively, just as the kind of significance which these things bear, but not by a relation with an external object. Sainsbury (2010 p. 300) says:

> Some thoughts are externally singular: there is an object which the thought is about. Some thoughts which are not externally singular are internally singular: although there is no object the thought is about, it recruits resources of a kind appropriate to external singularity.[2]

[2]On Sainsbury's model, all externally singular thoughts are also internally singular.

Frege (1948 p. 27) makes a similar move in distinguishing *reference* (i.e. my 'external content') from *sense*:

> It is natural, now, to think of there being connected with a sign (name, combination of words, letter), besides that to which the sign refers, which may be called the referent of the sign, also what I would like to call the sense of the sign

Many philosophers of perception have also written of how a perceptual experience 'seems' or 'appears' to a subject, regardless of whether this has any connection with real things in the world.

So there are two kinds of significance, external and internal, and it is widely agreed that both occur in a variety of phenomena. Now a third notion: content. Sometimes I will write 'content *simpliciter*' to distinguish this from the previous two. It is widely believed that some form of either external or internal content is privileged—that it is the real or primary content. This is sometimes called 'intentional content', with language, thought and perception as 'intentional phenomena', hence the title of this thesis, but I find this confusing and will write just of 'content'. The claim of Unity is that the nature of content—whether it is external or internal—is the same in language, perception and thought.

## 1.2 The road ahead

The first goal of this thesis is to provide a good reason to believe Unity. In Chapter 2 I survey the literature on questions of content in language, perception and thought since Frege. Within each area the same dialectic plays out. There are two positions available: that content is just external content, or that it is just internal content. The former view is that of the Millians, the naïve realists and the *de re* theorists of singular thought. The latter is that of the descriptivists, Fregean sense theorists (except McDowell), representationalists, and the *de dicto* theorists of singular thought. I demonstrate that the same types of puzzles and cases are raised by the same type of theorists, against the same views, in each domain. I divide these into four: *failure cases*, *Frege's puzzle*, *alternative-satisfaction cases* and *non-satisfaction cases*.

I argue that in an important sense this exhausts what can be said about content. Given this, and the prevalence of this dialectic across the literature, I give a 'methodological' argument for Unity.

My second goal is to provide an account of what content is, given that Unity is true, but this must wait until Chapter 4. As the discussion of Chapter 2 may

show, it is not entirely clear what kinds of things are supposed to *have* content. Even if it is clear, the obvious diversity of reasonable candidates might put Unity in doubt. Names, thoughts, perceptual experiences and related phenomena are very different! In Chapter 3 I give an account of the bearers of content. I introduce the neutral term 'refeme' for whatever it is that has the kind of content under discussion. So there are linguistic refemes, perceptual refemes and cognitive refemes. Against various received views, I argue that linguistic refemes are utterances of terms (in particular of names); that cognitive refemes are tokens of concepts or mental files; and that perceptual tokens are 'object files'. These turn out not be very diverse at all, so the worrying disanalogies do not arise.

In Chapter 4 I come to the main question: the nature of content. I argue that in order to make sense of the pervasive dialectic of the four puzzles with which I began, content must include both internal and external aspects. Theories of this kind have been held before, and have suffered on two fronts. First, it has often been unclear what the internal content is supposed to *be*. Frege, for instance, distinguishes two levels of meaning, but never offers an account of the nature of sense. I claim that internal content is a set of properties judged by the relevant agent to hold of the external content (if any); I call this aspect of content *ascription*. Second, it has been widely assumed that on any such picture, what internal content something has must determine what external content it has. There are clear counterexamples to this, which are held as arguments against such views. Frege held that sense determines reference, and Kripke gave straightforward objections. I argue that ascription need not determine reference at all, and comment on the shortcomings of the arguments that have been made to this effect. After all this, I offer solutions to the four puzzles with which I began. I redescribe the cases from across the literature in my own terms, and describe which refemes have which content in each case.

If ascription does not determine reference, what does? Come to that, what determines ascription? In Chapter 5 I describe a general view of which refemes have which content. Since I have argued that ascription does not determine reference, I offer separate accounts of the determination of reference and ascription. Both accounts are inspired by the causal theories of name-reference developed by Kripke, Devitt, Chastain and Sainsbury. From these theories I take a two-part structure: that some refemes 'inherit' their content by a relation with other refemes, and that some get their content 'primitively'. I apply this model to both reference and ascription.

The principle advantage of this structure is that a refeme of one kind can inherit its content (ascription, reference or both) from a refeme of a *different*

kind. An utterance of a name can refer in virtue of its relation with a cognitive token, which in turn refers in virtue of its relation to a perceptual object file. As I complain above, it is a common idea that relations between language, thought, perception and world are all interdependent in this way, but it is rarely worked out in detail. I hope that this account goes some way towards doing so.

This affords a second set of solutions to the original puzzles. The first solutions redescribed the puzzling cases in terms of the content and refemes of my theory. But accepting this, we might still wonder why situations of that kind occur. A theory of determination will tell us—I offer accounts of how failure cases, Frege's puzzle, alternative- and non-satisfaction cases arise in the first place.

## 1.3 The place of the project

It may be helpful to describe how my project fits into the wider literature. My two major goals bookend the programmes of research which inspired this thesis in the first place: analytic approaches to language and mind with their roots in Russell and Frege, running through Kripke, Devitt, and Evans and eventually to thinkers like Martin, Crane and Sainsbury. As I describe in Chapter 2, a lot of this work takes something like Unity for granted, though relies more or less heavily upon it.

The first major innovation of this thesis is that I present an argument for Unity. I am not aware of any other explicit argument for this claim, though it is of great importance in the existing work I discuss (Sainsbury, Evans, Devitt, Kripke, etc. and their commentators). Thus Chapter 2's defence of Unity can be taken as retrospectively legitimating analytic approaches bridging language and mind over the last 120 years.

Conversely Chapter 3 can be taken as tidying up one of their problems. If Unity is true, then long-standing assumptions about what *has* this kind of significance must be reevaluated. It has been generally assumed that the bearers of linguistic content are words or names,[3] those of mental content are thoughts, and those of perceptual content experiences.

First I show briefly that these three assumptions together are problematic for any view on which Unity is true, including those of the authors named above. Then I offer a new view of their nature: I say linguistic refemes are utterances, cognitive refemes are tokens of concepts in thought, and perceptual refemes are object files.

[3]Devitt (1981) is a notable exception.

My view differs from Kripke's, Evans' and Sainsbury's on language, as they all take names to be linguistic refemes; Devitt's view roughly agrees with mine. I differ from all four on cognition; Sainsbury takes concepts to be cognitive refemes (2005; 2010) and the other three take them to be thoughts. My view of the nature of perceptual refemes is entirely new. Even those who have written on perception, tacitly assuming Unity, like Martin and Schellenberg, have not discussed the question directly at all. When necessary to articulate their views, they assume that experiences are perceptual refemes.[4] Those for whom perception plays a role but is not the main focus, like Sainsbury, Evans and Devitt, have assumed the same. I am the first to argue positively for a specific view, and the first to hold the view that I argue for.

Chapter 3 is really just ground clearing for the positive views I present in Chapter 4, on the nature of content, and Chapter 5, on how content is determined. My positive views are descendants of a line of thinking with its roots in Kripke and Devitt, and especially in Evans and Sainsbury. Their views differ, but share the idea of a complex network of interrelated bearers of content, grounded in the world by relations of causation.

There are several important differences in the contents of our theories. My theory of content includes both internal and external content, but I argue explicitly that even on such a theory, internal content need not determine external content, contra previous views. Several others have argued for views of content involving two aspects—most famously Frege's sense and reference—but have assumed that internal content determines external content, and thus end up giving more weight to the former. My theory affords both an equal place. My theory of determination, though similar to traditional causal theories, can accommodate cases of non-causal inheritance (such as in writing), and is explicitly formulated to accommodate cases of indeterminacy. But the most important improvements are methodological. Beyond claiming that I am right and they are wrong, my theories are innovative in two ways.

First, my theories do not assume the primacy of philosophy of mind over language, or vice versa as they did. There is a long-running debate over whether, in the final analysis, the workings of the mind are to be explained in terms of the workings of language, or those of language explained by those of the mind. Dummett (1993 Chapter 13) offers some excellent commentary. Accounts of the former kind struggle to explain the nature of what language expresses or encodes; Dummett attributes this position to Frege and Davidson. Accounts of the latter kind struggle to explain the public or social nature of linguistic

[4]Martin (2002a) writes of the 'representational properties of perceptual experience'; Schellenberg (2016) compares the properties of 'three distinct, consecutive experiences'.

meaning; here Dummett mentions Evans, Peacocke and Chisholm. Dummet also spends a lot of time discussing 'idiolect'-driven theories, like Quine's, which try to strike a balance between the two.

With respect to the phenomena I am interested in, I remain entirely neutral on this question. Content *simpliciter*, and the external and internal content which I claim make it up, can both inhere in linguistic items and mental ones (where 'mental' includes perceptual). Both can refer to the objects in the world, and in this regard neither is parasitic on the other. On my account, bearing content emerges as a functional role which can be realised by either mental phenomena (cognition, perception) or linguistic ones (utterances, inscriptions). In theory, this even leaves my view open to the possibility of distinct kinds of content bearer which are neither linguistic nor mental, though I have not discussed this.

Dummett also says that all these thinkers accept one of the 'principal steps' of analytic philosophy: that whole sentences are explanatorily prior to their constituent terms (linguistic formulation), or whole thoughts prior to individual concepts (cognitive formulation). I reject this principle in Chapter 3.

Second, my theories are complete in a way that earlier ones were not. The most sophisticated accounts of the nature of content for which Unity is true are offered by Evans (1973; 1982) and Sainsbury (2005). Both are important influences on my work. But both rely on important, unanalysed and unremarked notions from which much of their utility is ultimately derived.

Evans (1973; 1982) relies on a notion of 'information' which mental states and linguistic items can 'have'. Information has an 'object' from which it is derived or 'of', and satisfiable 'content' which need not be satisfied by its object (1982 p. 128). One piece of information may have multiple sources; its object is whichever is 'dominant' (1973 p. 201). The content can be specified, 'with some loss' by an open sentence of first order logic, with free variables standing for objects (1982 p. 125). Information passing between different states and agents in complex networks is used to explain the nature and variety of reference, as well as communication and indexicals. But Evans never says what information is supposed to be, what its dynamics are, or what the content really is (i.e. without the loss of the logical interpretation). The external content of a state or token is ultimately just the origin of its information, and the internal content is the 'content' of the information (in Evans' sense of that term). Since these are not ultimately defined, Evans' theory is a careful exercise in accurately naming the problem.[5]

[5]A similar criticism has been made of Frege's lack of a theory of senses. In fairness to Evans, his work is illuminating on many related issues (many of which I do not discuss—see §1.4) and I suspect his failure to provide a solid foundation was due more to his illness and untimely death than any lack of insight or ability.

Though he is mostly interested in questions of reference, Sainsbury (2005) also recognises the need for some kind of internal content, and also relies on a notion of information for states and agents, though he says even less than Evans:

> Proper names help us store, sort and label information. The storing and sorting have a social dimension: memories are pooled over the whole. Proper names can be used to transmit knowledge. Belonging to a practice of using a name NN, and so interpreting uses of the name by others, can help one attain NN-related knowledge. This places a constraint on the transmission relation.
>
> (Sainsbury 2005 p. 111)

Sainsbury then uses the term 'information', without further elaboration. He also has a notion of object-related intention. He says:

> An object-related intention (I stipulate) is a mental state properly described along these lines: there is an object x such that the subject intends that … *x* … . [ellipses in original]
>
> (Sainsbury 2005 p. 106)

Sainsbury then describes various views of how the relation with objects might work (it might involve the object itself, might be representational, etc.). But he continues immediately:

> I have no need to choose between these views. I assume [only] that for an object to meet the relevant condition it must have played a suitable causal role in the origin of the object-related mental state.
>
> (Sainsbury 2005 p. 107)

As in Evans, certain states or actions can bear informational relations with each other, and can include object-related intentions, which ultimately explain relations of reference through causal networks. But since Sainsbury never explicitly describes how this works, as in Evans, Sainsbury's account doesn't deliver answers we might reasonably expect—the relation with an object, by which external content is achieved, is a primitive within the notions of 'object related intention' and 'information'.

Not only are Evans' and Sainsbury's ideas less specific than we might like, but this makes them difficult to scrutinise, because it is not clear how exactly they will analyse *recherché* cases which could be held up as counterexamples.

Like Evans and Sainsbury, I appeal in Chapter 4 to a non-referential aspect of content, which plays a role in the dynamics of content and reference in extended

(generally causal) networks. I call this aspect of content *ascription*. But I have tried to make my theory of ascription as clear as possible. In particular, I give a definition of what ascription is, and an explicit formulation of the role it plays in my theory, in terms of antecedent notions like belief, and describe exactly how internal content *works*. I have not been able to define everything, but where I leave things primitive I have noted this explicitly. I also treat explicitly the ascription of properties which themselves involve other objects, like the property of *being in France*. On the face of it, my theory strictly distinguishes questions of object-relations (reference) and properties (ascription). But some properties can involve objects, and two separate accounts will not cover them. As far as I am aware, I am the first to consider this problem. In this sense, I take my theory of content to be more 'complete' than theirs.

The same is true of their theories of determination—i.e. of which bearers have what content. Sainsbury and Evans both appeal to these notions in their theories of determination, and similar criticisms apply just as well there too, but there is a further incompleteness in their theories and Devitt's. All three develop varieties of a causal or historical theory of reference for names—roughly, that an utterance of a name refers to what it does because it is part of a chain or network of utterances, which is eventually grounded in a 'baptism' or initial event linking the chain (and thus the name, and thus the utterance) with the referent. So there are two parts to the theory: of how utterances are related to each other, and of how the chain is initially grounded in the object.[6]

Sainsbury, Evans and Devitt all provide accounts of both, but in all three cases the latter account does not adequately describe how the chain comes to be related to a particular object in the first place. Evans and Sainsbury's accounts both ultimately rely on their notions of information and object-related intention. For Sainsbury, the grounding is a baptism event, and the object baptised is whichever object the speaker's object-related intention is directed upon. (2005 Chapter 3, Section 5) For Evans the same role is played by the acquaintance that 'producers' in a name-using practice have with the referent. (1982 p. 376)

But since Sainsbury does not provide a theory of object-related intention, nor Evans one of acquaintance, the relationship which actually grounds the chain remains mysterious—the name-object relation is replaced with the baptism or producer-object one, and this is then replaced with the agent-object intentional or acquaintance one. But if the analysis ends here we have still not learned how the object gets to be part of the story.

[6]Though Evans (esp. 1973) sharply criticises causal theories like Kripke's picture, the producer/consumer model he eventually suggests does have this minimal structure.

Devitt's theory of grounding is better but has the same problem. For him, all reference is ultimately determined by relation to a 'grounding' event, which 'consists in the subject coming to have grounding thoughts about [the] object as a result of perceiving [that] object' where:

> A grounding thought about an object includes a mental representation of that object brought about by an act of perception. … Underlying [such representations] in those thoughts is a d-chain grounded in that object in virtue of which they designate the object.
>
> (Devitt 1981 p. 133)

But all this gets us no further. An utterance of a name refers to its referent because it is related to a grounding; the grounding involves a grounding thought; the grounding thought is related to the referent because *it* is related to a perception of the referent; the perception is related to the object by a D-chain (roughly: a causal chain); finally the thought designates its object 'in virtue of' this D-chain. But what does this mean? How does the chain 'underlie' the representation? Clearly the answer will have something to do with causation, but just exchanging one object-relation for another will not in the end tell us anything about how such object-relations are achieved.

In all three cases then, the question of reference is deferred to some other relation between intentional phenomenon and object, but this is just reference by another name, and is no less mysterious for it. To be clear, my criticism is not that these theories are not informative—which they are—it is that they are not sufficiently informative on this particular question.

I have tried to present a more complete theory. I also adopt the two-part structure common to these theories. But I offer a theory of the initial grounding, which I call 'primitive' reference, entirely in non-referential terms. In this sense, I take my theory of determination too to be more 'complete' than theirs. Like most philosophical theories, my theories are probably wrong, but at least it will be possible to know exactly why.

## 1.4 The scope of the project

This is a narrow project, and before beginning it may be worth making clear some of its limitations.

First, my theory is not a complete account of language, or even of meaning, or of cognition or perception. I am concerned only with the aspects of these which play the role of content, described above.

This is particularly important as some aspects of my account are quite cognitively demanding—reference and ascription both require the capacity for certain kinds of judgement. Some might worry that there are creatures which clearly perceive, cognise or communicate, but do not rise to this standard. This is entirely possible (indeed it is likely) but it is not a problem, because my theory is not supposed to cover all perceptive, cognitive or linguistic activity.

Second, the term 'content' has been used in many different ways, across language, thought and perception, often in conjunction with or reliance upon the topics of this thesis. But these are not the topic here. I am not concerned with phenomenological content, with truth or compositionality in sentences or extended works like novels or plays.

Third, discussions within philosophy of language of the issues I *am* concerned with are often presented alongside other theories of related phenomena, as if they were part of single package (which sometimes they are). These include theories of modality (both metaphysical and linguistic), rigidity, anaphora, quantification, indexicality, tense and propositional attitude contexts in language. I do not touch on any of these.

Three of these related phenomena deserve particular mention. I have not written anything about kinds. The work I engage with discusses the content of singular terms, singular thoughts, and objective perceptual experiences. These are phenomena which paradigmatically relate agents to individual particulars. In other work, these discussions are often presented alongside parallel views of analogous relations between agents and kinds or types (names and kind terms, singular and general thoughts, objective perception and the 'mere' sensation of qualities). This comparison is often used to support or undermine the original view. This may be a fruitful direction, but it is not my topic.

I have also not written anything about communication. I present a theory of the unified nature and determination of the content of object-relating phenomena like names and thoughts. A widely held slogan is that communication requires coordination on content; two agents communicate just when one gets the other to have a certain content (usually whatever content the first agent started off with). Supposing that this slogan is true, compatibility with such accounts is a constraint on the theory I give. But it is a much weaker constraint than might first appear. The 'content' I discuss is *only* that of names and singular thoughts, in so far as they relate to particular objects. The content involved in communication, if the slogan is right, involves much more—kinds, functions, propositional attitudes, modality, and the relation of all these to certain kinds of thought and language etc.. Not only this, but to explain communication, these must be supplemented with theories of compositionality in language, of linguistic force, and finally

of inter-agent coordination. Between the number of different phenomena to analyse, and the number of different theories available for each, the boundaries the slogan draws around my theory are very wide. I would be shocked to discover that my theory is either outside them, or constrained to such a specific area that well-respected theories of communication constitute objections to it.

Lastly, I do not provide a theory of identity claims. In particular, I do not discuss the problem of intentional identity. Take the sentence:

> Hob thinks a witch has blighted Bob's mare, and Nob wonders whether she (the same witch) killed Cob's sow.
>
> (Geach 1967 p. 628)

The sentence is unobjectionable. Clearly there is some connection between the mental states it attributes to Hob and to Nob. The puzzle is to give an account of the sentence's semantics which makes sense of this connection, but without undue ontological commitments (and in practice, which extends naturally to other very similar claims with slightly different commitments—Edelberg (1986) makes much of this). As Edelberg's treatment makes clear, the real topic here is the correct semantic analysis of sentences making complex belief ascriptions. Whatever the correct theory, it is clearly to be explained in terms of a relation between separate tokens—either between linguistic tokens in the sentence ('a witch', 'she'), or mental tokens in any situation which would make the sentence true (Hob's belief and Nob's). I provide two accounts of such relations in Chapter 5: of the inheritance of reference and of ascription. I am sure these resources would be helpful in providing a full theory of intentional identity, but as with communication, a full theory requires much more: accounts of the semantics of propositional attitude verbs, of anaphora, of how these are related, and perhaps more. What I say is relevant, but since I do not treat these topics, I do not provide a full theory.

Fourth, though I comment briefly on demonstratives and definite descriptions, I have not presented a full account. This would be an interesting extension to my theory, but they are not the main concern. My theory of determination could be extended too. The only other theorists to engage seriously with something like inheritance but without causation are those who hold 'anaphoric' deflationary theories (e.g. Brandom 1998). On these views, reference and anaphora are closely related, so my theory might cover anything they do.

Finally, the most obvious omission is any treatment of the 'content' of names, concepts or mental files. I argue in Chapter 3 that 'long-lived' entities like these, which can persist between occasions of use, do not have content properly so-called. But many theorists do talk about the content of names and concepts. I

am sure that a suitable (suitably different) notion of content could be recovered for these kinds of thing by coupling my account with a way of aggregating or collecting contents of many tokens; the content of a name would be some kind of aggregate of the contents of its utterances and inscriptions.[7]

This would bring my project full circle. Among those who discuss the content of names and concepts are Evans and Sainsbury. I have framed this project in terms of theirs: first with Chapter 2 justifying it, and then with Chapters 3 to 5 tying up its loose ends. A developed notion of content for names and concepts would explain not only this, but also how the useful work already done by Evans and Sainsbury could be recovered in terms of the details I present. Sadly, this was beyond the scope of this thesis.

The theory I produce, including my argument for Unity, is conjectural. It is an attempt to provide *a* good way of making sense of some widely held assumptions, of sufficient comprehension and detail to enable further work. I have tried to a present a more explicit account than some who have gone before, and occasionally this has been at the expense of more rigorous argument or consideration of alternatives. This is a difficult balance to strike; I have done my best.

[7]These rather suggestive remarks will make more sense in the light of my eventual account.

# Chapter 2

# History and Motivations

Recall the claim of Unity:

**Unity** *The nature of the significance of names, perceptual experiences and singular thoughts is the same.*

Unity is vague and poorly specified. This is deliberate, and for several reasons. First, it is often not clear what exactly various philosophers have believed (often not clear to readers or even the authors themselves)—the versions of Unity they believe really *are* just vague and poorly specified. The fact that Unity is implicit or tacit in most writing only makes this worse. Second, to the extent that it *is* clear, different philosophers have definitely believed different versions of Unity, and with different attendant assumptions, depending on their theoretical background (compare: two philosophers might both be structuralists about numbers, though their views actually differ in the details). Finally, these people have often taken themselves to be in dialogue with one another, notwithstanding their actual theoretical differences. So I take it that there is a general, though not well specified, claim, shared by many, expressed by Unity.

For example, Bach (2010) is a discussion of the relation of singular thoughts to objects, but the paper begins with an example of the relation between *perceptual experiences* and their objects. In Section 7 Bach says 'For now I just wanted to have a vivid analogy to the problem of singular thought, although in relevant respects they are pretty much the same problem'. Bach never makes explicit what he takes the relation between the two problems to be, but the general insight is clear enough. He is typical in both his assent to Unity and his lack of clarity; such claims are widely held, though they have often gone unmotivated or unrecognised, and thus unspecified.

The purpose of this chapter is to develop a good reason to believe Unity, a reason which will turn out to be mostly historical. I make no efforts here to make Unity more precise. Instead, the following chapters begin from the view that

something roughly expressible as above is true, and that what I say in this chapter provides good reason to believe it. With these claims in hand, the later chapters say more precisely how Unity should be interpreted if it is taken to be true for those reasons; Chapter 3 argues for how 'names, perceptual experiences and singular thoughts' should be understood, Chapter 4 for what 'significance' is.

Philosophy often proceeds by considering puzzling cases. In this chapter, I argue that discussions of 'content' in any of these three areas since the work of Frege have been largely concerned with four kinds of puzzle: *failure cases* (§2.1), *Frege's Puzzle* (§2.2), *alternative-satisfaction cases* (§2.3) and *non-satisfaction cases* (§2.4); these labels will be explained further as we go along.

Each puzzle is a type of case; a schematic kind of situation deployed in thought-experiments for or against certain views. For each, I show that cases of this kind have been *used* in the literature of all three domains (language, perception, singular thought) and moreover that they have been used to motivate or contradict similar kinds of view. The important points to take away are that the *same* kinds of puzzling cases occur in the literature on language, perception and thought, and that they play the *same* dialectical role in each area.

The strands come together in §2.5 with my argument that Unity is true. First I present a brief epistemological argument that the obvious similarity in the puzzles makes it seem likely they instantiate a common kind. I don't rely on this argument, but I note it because I believe this is roughly what has actually brought many philosophers to believe Unity.

I then offer a methodological argument: given the minimal notions of internal and external content, and some widely-accepted assumptions, the dialectic force of the puzzles considered in §§2.1 to 2.4 *exhausts* what can be said for or against prevalent theories of content. Since this exhaustion is the same for language, perception and thought—and can be shown to be so without appeal to the particular characteristics of any of these—it is very likely because language, perception and thought share some common nature. That is, Unity is true.

As far as I am aware, I am the first to give an explicit argument in favour of Unity, though many have relied on it.

This argument relies heavily on the methodological utility of certain kinds of case. This will be a theme throughout the thesis. Accordingly, this chapter also serves to give clear characterisations of these four puzzles, which will be useful later on.

## 2.1 Failure cases

Recall the distinction I made in the introduction between internal and external content: external content is the significance a thing has by its connection with a worldly object (or for some theorists, its merely *apparent* connection). If the thing has a referent, generally the referent is thought to be the external content. Internal content is typically characterised negatively, as some kind of significance which is *not* constituted by the things' connection with a worldly object (or at least not *merely* so constituted).

*Failure cases* are examples of tokens with no external content (or at best, some sort of empty or null external content) but which appear to have some kind of significance, similar to tokens which *do* have external content. Presented with such a case, we are invited to share the intuition that these tokens *do* have content (in the unified sense). Since they have no external content, these cases are supposed to demonstrate that content is not identical to external content. In response, various theories have been developed on which content is entirely internal, or where external content is determined by internal content. In this section, we will see that this dialectic plays out in exactly the same way across philosophical work on language, singular thought and perception.

Linguistic failure cases are examples of apparently meaningful 'empty names' (sometimes 'vacuous names', especially in older literature). These cases are generally deployed against some variety of Millian or direct reference theory. These theories claim that the meaning of name is a that name's referent.

Before the main discussion, two important points on names. First on names and utterances. Most of the views discussed below assume that it is names which bear content and reference, rather than utterances or tokens of those names. Though it is sometimes awkward to state the views in this way, I have followed the style of the works I am discussing. Second on ambiguity. Many of the cases I discuss are successful only if each name is treated as either referring to one thing, or failing to refer at all, though it seems many real names are ambiguous. There was an Aristotle who taught philosophy and there was another who sold ships; neither is the 'real' or unique referent of the name. Nonetheless many writers have found it useful to adopt this as a terminological convention or idealisation (e.g. Kripke 1980 p. 8; Sainsbury 2014 argues that this is no idealisation but actually true). Since the purpose of this chapter is to chart (patterns and trends within) the history of these discussions, I have found it most straightforward to continue in this way. My own discussion in the next chapter proceeds in a rather different way, and arrives at a view of the bearers of content and reference on which the first assumption is false, and therefore the second is inapplicable, so there they are dropped.

Russell is an early example of discussion of empty names.[8] (1905; 1919; 1956b) Russell was attracted to a roughly Millian view of proper names. But he noticed that we seem to use proper names for things which do not or may not exist. His favoured example is that we might ask whether Homer existed. (1919 p. 179; elsewhere Russell also discusses 'Pegasus', 'Apollo' and 'Hamlet'). Clearly if the question is to be intelligible it must be in question whether Homer existed. But if 'Homer' is a name in the Millian sense, then just the fact that we can use it as such indicates that Homer did exist, so the question is pointless. If it is not a name even though it is used as one, then the question is unintelligible. But in fact asking whether Homer existed is neither unintelligible nor pointless. So, Russell concludes, the significance of 'Homer' must go beyond its referent (or status as non-referring as the case may be). Russell was very changeable about what this actually means: in "On Denoting" (1905 p. 491) he briefly treats empty names like 'Apollo' and 'Hamlet' as of the same kind as denoting phrases like 'the round square', though he gives no explanation of which predicates their corresponding logical form will involve, and does not mention non-empty names; this is the earliest and least sophisticated view. Later he says that 'names we commonly use … are really abbreviations for descriptions' (1956a p. 201) so perhaps 'Homer' is a name, though not a Millian 'logically proper name', and the nature of such names is to abbreviate descriptions which eventually denote objects. Later still, he wrote 'Common words, even proper names, are usually really descriptions' (1970 p. 29) which seems to deny that 'Homer' is a name at all. But Russell's many solutions are not the concern here. Throughout, Russell consistently held that the intelligibility of apparent empty names poses a problem for a view like Millianism; the fact that he recognised this is all I rely on here.

Quine (1960) lists a number of similar challenges before saying:

> Finally, independently of all such technicalities, there is just something wrong about admitting that 'Pegasus' can ever have purely referential position in truths or falsehoods [i.e. contribute only its referent to the truth value of a sentence]; for the intuitive idea behind "purely referential position" was supposed to be that the term is used purely to specify its object, for the rest of the sentence to say something about (§30). Singular terms which, like 'Pegasus', lack their objects thus raise problems.
>
> (Quine 1960 Section 37; see also discussion of 'Pegasus' in Quine 1948)

[8] Historical discussions of theories of names generally begin with Mill and Frege, presenting Russell as a foil to one or both. (e.g. Sainsbury 2005 Chapter 1) Neither had much to say about empty names, and Frege's principal concern—with different but co-referring names—is covered in detail below.

Like Russell, Quine is concerned with a tension between the apparently intuitive thesis that names just mean their bearers, and cases of apparently meaningful names without bearers. Like Russell, Quine took this as a refutation of a purely referential theory of meaning for names. Quine thought that names should be 'reparsed' as general terms, each true of at most one thing (in this, Quine's theory is rather similar to Russell's descriptivism, thought the technical details are different).

This tension is brought out clearly by Grice's (1969) response to Quine. Grice argues thus: we would like to hold at least that 'Pegasus' has no referent (Grice's 'inclination 2') but also that there can be sentences using 'Pegasus' which are about Pegasus (inc. 3). If we then admit the laws of identity (inc. 7), universal instantiation, and existential generalisation (inc. 6), then from the plausibly true 'it is not the case that Pegasus flies' we may derive the clearly false 'Pegasus exists'. Grice's argument is a slightly different formulation of the same worry Russell and Quine had: it is attractive to hold that if some proper name is meaningful, then there must be a *thing* (the thing named) in virtue of which it is meaningful; 'Pegasus' is meaningful; so there is a thing, Pegasus. But clearly there is no Pegasus, so the view of the meaningfulness of proper names must be at fault. Grice's proposed solution is powerful and complex, but will not see much discussion in this thesis, because it relies heavily on notions of scope and syntax which are only applicable to language (perhaps even only formal language). Since the goal of the thesis is to describe a theory of content (and thus, a solution to these puzzles) which is equally applicable to language, perception and thought, solutions like Grice's are dead in the water.

After Kripke's (1980) attack on descriptivism (of which more below), there are few such responses to the problem of empty names and most philosophers turned to Millianism.[9] But this only makes the problem more pressing. If we are to hold Millianism about *all* proper names—not just some privileged subset as Russell does—then we must positively explain the apparent meaningfulness of names like 'Pegasus'—not just explain them away, as Russell does.

McDowell (1977) develops a sort of neo-Fregean Millianism, on which knowing that 'Hesperus' refers to Hesperus exhausts what it is to know the meaning of 'Hesperus' (this is the Millian part—the neo-Fregean part is how this is interpreted). McDowell develops this idea in the context of a Tarski-style truth theory, where for each name n, there is an axiom "n' refers to n'. These axioms play a part in the derivation of the theorems concerning truth. To know the meaning of a name, McDowell claims, it is sufficient to know the axiom governing it (much

[9]Lewis (1984) is an exception to the general anti-descriptivist trend, though he has little interest in empty names.

of the paper is taken up with a difficult discussion of what such knowlege is). But empty names pose a problem for this setup. On one hand, for an empty name e, there will be no axiom "e' refers to e'. On the other, it seems that speakers and hearers *do* know the meanings of names like 'Pegasus'. Equivalently, understanding the meaning of 'Pegasus' cannot be explained by knowing that 'Pegasus' refers to Pegasus, since there is no Pegasus that the term could be known to refer to. McDowell is aware of this issue and offers a response, but the important point here is that the problem is basically the same as above: McDowell wants to account for the meaning of names entirely in terms of reference, but finds it hard to explain apparently meaningful non-referring names.

Sainsbury (2005) takes a similar line.[10] The project of the entire book is essentially to recover the claim that:

> There is a single category of referring expressions, all of which deserve essentially the same kind of semantic treatment. Included in this category are both singular and plural referring expressions ('Aristotle', 'The Pleiades'), … and empty and non-empty referring expressions ('Vulcan', 'Neptune').
>
> (Sainsbury 2005 p. viii)

Sainsbury argues for an 'austere' theory of meaning like McDowell's, but also that this theory should be given in a negative free logic, thus recovering the intelligibility of empty names. Like Russell, Quine, McDowell and others, Sainsbury sees the attraction of a theory of names' meaning given only in terms of their referents, but also that this won't be sufficient to explain empty names.

For more modern treatments in a similar vein, see Searle (1969 Section 722, objection 3), Braun (1993), Salmon (1998), Everett (2003), Adams and Stecker (1994), and Cullison and Caplan (2010). A different and persisting line of thought approaches the problem of empty names by considering the use of names in fiction and pretence; see Inwagen (1997), Thomasson (2003), Sainsbury (2003c), Braun (2005), and Tiedke (2011).

Common to all these cases is the following dialectical move: although there may be theoretical virtues to a view on which the significance of a name is merely its referent, such theories are unable to explain the obvious intelligibility of (uses of) names with no referents. So such theories, at least without modification, must be wrong.

[10] *Reference Without Referents* is the eventual culmination of at least two distinct lines of thought in Sainsbury's earlier work: the development of an Austere theory of sense like McDowell's first developed in 2003d and the use of negative free logic to capture apparent truths about meaning and reference with empty terms, following Burge (1974), first discussed in 2003c.

We see a similar pattern in the philosophy of mind. The phrase 'singular thought' is used for thoughts which are in some sense about or related to a certain, specific object. There is no general consensus in the literature on any more specific definition[11] but a paradigm instance would be my thinking of my father that he is bald. What is important is that my thought is of some *specific* particular, my father, not just any old particular (or even any old particular which fulfils some certain condition). Singular thoughts are generally contrasted with 'general' thoughts, which are usually understood just as thoughts which are not singular thoughts.[12]

Cognitive failure cases are singular thoughts about non-existent entities.[13] Such cases are generally deployed against some variety of 'object involving' theory of what singular thoughts are. According to such theories, the content of a singular thought literally *includes* the object thought about, or at the very least is constitutively dependant upon it. Crane (2011b) calls this the 'orthodox' view. The problem for the orthodox view is that one can have thoughts about objects which do not exist, but which seem—to oneself and to philosophers considering this—to be of the same kind as the thoughts about existing objects. Such thoughts are often subjectively indistinguishable from thoughts which relate to real particulars, and they seem to play a similar role in our mental and social lives. In other words: there seem to be singular thoughts about objects which do not exist.

Notice the clear parallel with empty names. Millians believe that the content of a name is its referent; no referent, no content. But empty names appear to have content in spite of not having a referent. Orthodox theorists about singular thought believe that the content of a singular thought is (or constitutively depends upon) the object thought about; no object, no content. But singular thoughts about nonexistents like Pegasus or Vulcan appear to have content, despite there being no object.

Chisholm (1977 Section XI) is an early example. Chisholm doesn't use the term 'singular thought' but the problem is clear. His case is a round square: we can think of round squares, so it can be true that I am thinking of a round square. But round squares do not exist, so it is true that I am thinking of something

[11]Though some philosophers write as if there is, and they known what it is. See Jeshion (2010b) for a survey of recent work.

[12]These terms seem to have their origin in the distinction distinction between singular and general *terms*. Though often attributed to Quine (1960 e.g. p. 82, pp. 86–95), the distinction is clearly present in Locke (1975 Book 3, Chapter 1, Section 3).

[13]Or at least *putative* singular thoughts. According to some, aboutness requires existence, so thoughts which seem to be about non-existent entities are either not really *about* them, or are not really thoughts.

which does not exist. But, Chisholm asks, 'how can a truth be about an object which does not exist?'.[14]

McDowell (1982 p. 304) defines singular thoughts as those 'not … available to be thought or expressed if the relevant object, or objects, did not exist'.[15] So if one utters a sentence containing a non-referring singular term, in an attempt to express such a thought, one does *not* actually express any thought (because there *aren't* any such thoughts). But it seems that we *do* express genuine thoughts using non-denoting singular terms. McDowell concludes (with Russell) that these apparent singular terms were not really singular terms to begin with, and that the subjects who utter them are not really expressing any thoughts (though they and others might wrongly believe that they are, so that their actions appear intelligible).

McDowell does not seem particularly exercised by the apparent existence of empty singular thought. Segal (1989) takes the problem a bit more seriously and develops a careful criticism of McDowell's views:

> whenever [a theorist like McDowell] alleges that a subject, $S_1$, has an object-dependent singular thought, *o is F*, [about a worldly object] we can imagine that the subject has an identical twin $S_2$, who will undergo, as [McDowell] would have it, a failed attempt at a thought of the form *o is F*. Then I face [McDowell] with a challenge: make sense of $S_2$'s behaviour in a way that does not require attributing to her the thought *o is F*. If [McDowell] cannot to do this, so much the worse for [his theory]: such deluded subjects do act rationally, and we need to explain this. If [the theory] can make sense of $S_2$ without attributing to her the thought that *o is F*, then it will be possible to make sense of $S_1$ without attributing to her the thought that *o is F*. And if that is right, then it follows that it would be theoretically unmotivated to attribute the thought *o is F* to $S_1$. In fact it would be wrong. And this means that $S_1$ does not have the object-dependent thought *o is F* after all.
>
> (Segal 1989 p. 42, original italics)

[14]Chisholm moves comfortably between considering one's *thought about* something which does not exist, the truth *that one is thinking about* something which does not exist, and the locution 'the [square] one is thinking of' (as in 'the square you are thinking of is round, and therefore does not exist'). I take this to be part of Chisholm's solution to the problem of thought about non-existents, not a slide to a different puzzle altogether.

[15]McDowell's discussion is ostensibly an exegesis of Russell's (1905) views on singular thought. Russell never uses the phrase 'singular thought', and does not explicitly discuss mental phenomena at all until the penultimate paragraph of the paper. I take McDowell's comments to be best understood as his own views, developed by way of commentary on Russell's.

The contemporary debate on cognitive failure cases crystallises around the work of Tim Crane and Mark Sainsbury. Crane sees the attraction of the 'orthodox' view. It gives a simple explanation of what distinguishes singular thoughts from general thoughts: a singular thought 'ontologically depends on the existence of the particular object it is about' and a general thought does not. (2011b p. 24)[16] But Crane goes on:

> However, if there is genuinely singular thought about the non-existent, then this object-dependence thesis is false. And there do seem to be straightforward cases where a thinker's thought is aiming to refer to some particular thing which does not exist. … Those who think about Vulcan seem to be having thoughts that are aiming to refer to a particular object just as the thoughts about Neptune are.

As with empty names, the problem here is that if the content of singular thoughts about some X is to be explained entirely in terms of X itself, it will be impossible to explain the content of (apparent clear cases of) singular thoughts about Pegasus or other non-existent entities.

Crane's earliest position on this issue was a minimal account of the metaphysics of the objects of thought (2001). In later work (2011b; 2013) he takes a different tack and argues that singular thought should be distinguished from non-singular thought by its particular 'cognitive role'. The relevant role is a relation to a singular proposition, though some thought-episodes play this role without in fact relating to such a proposition. Zarepour (2018) calls this a 'psychologistic' account, because what matters is the nature of the thought itself, not its actual relation to the world.

Sainsbury says similarly:

> An appealing thought is that, in an intentional mental state, the mind is directed at an object: at, as one might say, an "intentional object." This seems unproblematic for cases like Bush's admiration for Cheney or Obama's careful thought about Iraq. Bush directs his mind at the object Cheney, Obama at the object Iraq. … It is problematic in cases like Bush's admiration for Superman or Obama's careful thought about Arcadia, since on intuitive views there is no such object as Superman or as Arcadia. These cases cannot be straightforwardly described as ones in which the minds of Bush or Obama are directed on objects.
>
> (Sainsbury 2010 pp. 302–3)

[16] Crane draws partly on material developed in earlier papers, see his work of 2001; 2011a,b; 2012.

Common to all these cases is the following dialectical move: although there may be theoretical virtues to a theory on which the content of a singular thought (or other cognitive element) is merely the thing thought about, such theories are unable to explain the obvious significance of thoughts where there is no corresponding worldly object. So such theories, at least without modification, must be wrong.

Finally the perceptual failure cases. Perceptual failure cases are objective hallucinations: perceptual experiences as if of a real object, though in fact there is no such object actually perceived.[17] Such cases are generally deployed against naïve realism (sometimes 'direct' realism), the theory that the content of any perceptual experience is just the thing or things perceived. The problem for this view is that there seem to be significant and contentful hallucinations, often of specific particulars, and it is impossible for the naïve theory to account for this. I might have a hallucination of my father (not just some man or other, but *my father* in particular), but the naïve theory cannot explain how I hallucinate *him*, because he isn't there to be perceived. Problems of this kind loom so large in philosophy of mind that they are simply called 'the problem of perception' (see e.g. Crane and French 2021 for an overview).

Price's (1961) discussion is one of the earliest and most nuanced. The most powerful version of his argument is 'The Phenomenological form of the Argument from illusion'. Price begins:

> An illusory sense-datum[18] of sight or touch is a sense-datum which is such that we tend to take it to be part of the surface of a material object, but if we take it so we are wrong.
>
> (Price 1961 p. 28)

Price gives some interesting examples which have been less considered in later literature: effects of perspective, reflection, refraction and after-images can all make it appear that there is a surface where there isn't. But the most powerful is hallucination, where 'the sense-datum fails to belong to any object in any way at all, though of course the [subject] still takes it to belong to one'. (p. 29) Price gives various examples, but the clearest is that 'when my hat is taken off my head, I still continue to feel it there for a time'. Then:

[17] I use 'perceive(d)' factively. To perceive is to experience veridically, so no hallucination is an instance of perception. I use 'perceptual experience' non-factively, so all hallucinations are perceptual experiences.

[18] Price uses the term 'sense-datum' as we would use 'direct object of perception'—a sense-datum for Price is just whatever is most immediately encountered in perception. Naïve Realism is thus the claim that 'the relation between a sense-datum and a material thing when the thing is present to the senses of the being who senses the sense-datum' is the same as the relation 'being part of the surface of'. (p. 26)

> The Phenomenological form of the Argument from illusion simply points to such facts as these and contends that they directly refute Naïve Realism. And it certainly does seem clear that there are a great many visual and tactual sense-data which cannot possibly be parts of the surface of material objects in any natural sense of the words 'surface' and 'part'. It follows that perceptual consciousness cannot always be a form of knowing, as the Naïve Realists say it is, for it is certainly sometimes erroneous.

The dialectic is now familiar: whatever attractions there may be in a theory which describes the content of perceptual experience entirely in terms of the objects actually perceived, there are clear cases of perceptual experiences, with significant content, but without any corresponding worldly object which could possibly explain it. In Price's hat case, the later experience of the hat cannot be explained be the felt presence of the actual hat, because the hat isn't there to be felt.

Martin (2002b) lays out the problem clearly:

> [it is widely assumed] that perceptual experiences form a common kind of mental state among cases of veridical perception, illusion and hallucination. The perceptual experience which one has when seeing a pig is of a kind which could have occurred were one not perceiving at all but having a visual hallucination indistinguishable from the sighting of the pig. On such an assumption, one may demand that whatever account one is to give of the experience one has when one veridically perceives, the same account must be applicable in some cases of hallucination. But, on [a naïve realist view], the objects of perception are aspects of the visual experience which actually have to exist when one has such an experience, so one could not have an instance of that kind of experience if the objects did not exist. One could not have such an experience in the case of hallucination. Therefore, given the assumption that perceptual experience forms a common kind across veridical perception, illusion and hallucination, the naïve realist view must simply be false.
>
> (Martin 2002b p. 393; see also Martin 2002a e.g. p. 20)

Other descriptions of the problem of perception are so similar to those above that detailed examination would be unilluminating. See for example Foster (2004 p. 5), Johnston (2004 p. 113), Crane (2005 p. 239; 2006 p. 133), Tye (2007 p. 594), and Sainsbury and Tye (2012 p. 151).

Common to all these cases is the following dialectical move: although there may be theoretical virtues to a theory on which the content of perceptual experiences is merely the worldly object perceived, such theories are unable to explain the obvious content of experiences where there is no perception of worldly objects. So such theories, at least without modification, must be wrong.

Martin's writing brings out a useful contrast between presentations of failure cases in perception and in language or cognition. The dialectic in work on perception often begins with the assumption (to be eventually challenged) that hallucinations and veridical perceptions are of the same basic kind, and that therefore the same theoretical explanation must be given for the content of each. This is often called the 'common kind claim'. Then we come to the suggestion that the content of veridical perception might be explained in terms of the objects perceived, then the observation that the content of hallucinations cannot be. Martin and others have thus concluded that we should reject the common kind claim (this is *disjunctivism*).

Presentations of failure cases about language and thought often go differently: it is tacitly assumed by almost everyone that empty and non-empty names form a common kind, and similarly for singular thoughts about existent and non-existent things. The next two steps are the same as for perception: it is attractive to account for the content of names/singular thoughts entirely in terms of their respective worldly objects, but the existence of failure cases (empty names, thoughts about non-existents) makes this clearly infeasible. Russell, Quine and others have thus concluded that we should reject the analysis in terms of objects (and turn instead to something like descriptivism).

But the only part of either presentation which is important for my survey is what is common to both: that the existence of failure cases is philosophically important, and that this is so because they provide prima facie counterexamples to Millianism, the 'orthodox' view of singular thought and naïve realism, all of which analyse content in terms of the worldly objects of their respective phenomena.

It is true that historically, philosophers of language and cognition have been more likely to resolve this difficulty by rejecting the analysis in terms of objects (retaining the tacit common kind claim) whereas philosophers of perception have been more likely to reject the common kind claim (retaining the analysis in terms of objects). But this has no bearing upon my claim here, which is just that the same kind of problematic case shows up in each domain, and is problematic in each for the same kind of reason. It is also true that these tendencies are neither necessary nor universal: Tiedke (2011) presents a disjunctive analysis of proper names, McDowell (1982) rejects the possibility of singular thought

about non-existents, and certain intentionalist or representationalist approaches to perception retain the common kind claim while rejecting the object-based analysis (more on this below).

But two general points stand. First, there is a general structure of philosophically interesting case that has been marshalled in discussions of language, cognition and perception. The structure is that there are apparently significant instances of some phenomenon, which usually obtains its significance by relation to some worldly object, but where there is no such object. Second, that this kind of case has been marshalled *against* the same general kind of theory in each domain—theories which claim that the significance of any instance of the phenomenon is *always* identical with (or constitutively dependant upon) its associated object. I call such cases in general *failure cases*, because the instance of the phenomenon *fails* to relate to the world in the way it is supposed to.

## 2.2 Frege's puzzle

The next kind of case to consider is Frege's puzzle. Frege (1948) introduces cases of two, differing tokens, with identical external content but differing internal content. Suppose there are two non-empty names, 'a' and 'b' and consider the following two sentences:[19]

**Trivial** ⌜*a is identical to a*⌝

**Informative** ⌜*a is identical to b*⌝

Sentences of the form of Trivial are obviously true, because the same sign is simply used twice.[20] No knowledge of the meaning or reference of 'a' is required to see this. Suppose that Informative is also true. This is not similarly obvious; Frege says that the two sentences have different 'cognitive values' (ibid. pp. 25–26) and that a 'special act of recognition' is needed to see that statements like Informative are true, unlike statements like Trivial (Frege 1980 p. 152). The puzzle is to say why. There is a sense in which, if the two statements are both true, they must mean the same thing, since they both concern just the one object. But if they mean the same thing, would we not expect them to have the same 'cognitive value'? Clearly, a more careful notion of meaning (or cognitive value) is needed.

[19]Frege (1948 pp. 25–26) presents these as $a = a$ and $a = b$, but I have tried to stick to natural English.

[20]If we assume names are individuated merely by their graphical form, or even by form and reference, then there will be apparent counterexamples such as Kripke's (1979) 'Paderewski' case. I argue later that such counterexamples are merely apparent.

Appeals to Frege's puzzle, like appeals to failure cases, have been presented as evidence against views like Millianism. The argument goes that each name ('a' or 'b'), as used in a sentence, contributes its content to the meaning (the content) of the sentence overall.[21] Since the two sentences differ *only* in one instance of a name, if the two names had the same content, the resulting sentences would have the same content too, and thus the same cognitive value. They do not, so Frege concludes the names did not have the same content after all (Frege does not use the term 'content'). But *ex hypothesi*, the two names have identical *external* content—they share a referent—so content cannot be identical to external content. That is, a view like Millianism must be wrong.

Frege's alternative was a theory of 'senses'. The content of each name is not (primarily) its referent but its sense. A sense is a public entity, which can be cognitively grasped by a speaker or thinker. To every meaningful term there corresponds a sense, and each referring term's referent is determined by its sense. The puzzle is solved by the claim that two different names might have different senses, but these two senses determine the same referent. Trivial is obviously true because it only contains one name, so the same sense is grasped twice. Informative is not because it contains two names, with differing senses.

The problem is powerful, and widely studied. Within the philosophy of language and cognition, Frege's original description of the puzzle remains the standard.

The most sustained defence of Millianism in the face of this problem is Salmon's (1986). Salmon denies that there is any difference between the information encoded by Trivial and Informative—both just encode the proposition that Venus is Venus, since all the names in either statement refer to that one thing. This proposition can appear in various 'guises' though.[22] We might thus fail to recognise that one proposition is encoded by both sentences, and wrongly believe that in believing what is encoded by the second, we believe something *new* (not encoded by the first). In fact though, we learn the same things from Informative and Trivial (indeed we learn nothing, since anyone who knows of Venus knows that Venus *is* Venus).

In terms of the above description: Salmon denies that the cognitive value of the sentences actually differs, but admits that it might *appear* to an agent that it does.

[21]This is a simple principle of compositionality. One solution to Frege's puzzle is to simply reject this. Such solutions are beyond my concern here and will not be discussed.

[22]This is a slight simplification. Salmon does write in terms of propositional appearances and guises, though his professed view is that only *objects* really have guises. Propositions only have guises in some derivative sense, when their constituent objects do.

Soames (2002) is also impressed by Frege's puzzle. Considering two sentences which differ only in co-referring names, he says:

> Some speakers might be competent users of the names without realizing that they are coreferential. Presumably, they would not judge [the two] sentences to mean the same thing. Even among speakers who know the names to be coreferential, many might recognize that the sentences often would be used to communicate different information and, on that basis, judge the paired sentences to differ in meaning. In the case of [a sentence like Informative] and [one like Trivial] this intuition could be expected to be very robust, since the information carried by an utterance of [Informative] would nearly always be different from the information carried by an utterance of [Trivial].
>
> Many regard this as decisive evidence against [millianism].
>
> (Soames 2002 p. 68)

Soames' solution is pragmatic. He distinguishes the information communicated by one or another utterance of a sentence (which may vary from utterance to utterance, depending on context, the speaker's and hearer's background knowledge, etc.) from what the sentence *means*, which is context-invariant. The only candidate for meaning which *is* context-invariant is the Russellian proposition associated with the sentence. And this is determined by the constituent parts of the sentence, and their associated, context-invariant semantics; the context-invariant contribution of a name is its referent.

McDowell puts the problem like this: 'how could someone possess knowledge sufficient to understand a sentence like 'Hesperus is Phosphorus' without thereby knowing already that the sentence is true?'. (1977 p. 163) The concern is that if the speaker does understand the sentence, surely they must understand each of the two names. In the wider framework McDowell is developing, to understand a name is know its object (McDowell distinguishes knowledge of truths from knowledge of objects). But the names name the same object, so it seems that if one understands the sentence (thus understanding both names) one already has the knowledge that the sentence expresses, so the sentence could not be informative. But the sentence *is* informative—what has gone wrong?

His solution has many layers, but the main insight is that knowledge of an object is not the same as knowledge of a truth. One might well know which *object* 'Hesperus' refers to (Venus), and know which object 'Phosphorous' refers to (Venus again) but fail to know the *truth* 'Hesperus is Phosphorous' (or equivalently, in McDowell's truth-theoretic framework, that "Phosphorus' denotes

(stands for) Hesperus'). To know the sense of a name just is to know its referent (qua object), but one might fail to know various truths about the referent, even though one knows *it*. McDowell thus recovers a notion of sense powerful enough to solve the puzzle, but without having to posit a theory of senses themselves. Sainsbury is much impressed by this 'austere' theory of sense, and endorses a similar view (see e.g. Sainsbury 2005 Chapter 1, section 5).

So the puzzle has been approached in different ways, but common to all is the following initial dialectic: there are pairs of sentences, utterances of which would be associated with systematically different behaviour and beliefs in non-philosophical competent users of the language. So, the sentences must differ in meaning. Since their meaning is determined by the meanings of their parts, and their only difference is the exchange of one name for another, these two names must differ in meaning. But the names co-refer. So, meaning (i.e. content) cannot be the same as reference.

Frege's puzzle has traditionally been understood as a problem for theories of language, cognition, or related notions like belief (as in Salmon and McDowell's epistemological solutions). In particular, for theories like Millianism which theorise content in terms of the actual object to which a name or thought relates. We saw that failure cases are problematic for such views of language, cognition and perception. It is no surprise then that Frege's puzzle has been extended to perception as well. Millar is very clear:

> According to [views such as naïve realism][23] the contents of perceptual experiences include singular propositions partly composed of the particular physical object(s) a given experience is about or of. [Any such] view faces well-known difficulties accommodating hallucinations; I maintain that there is also an analogue of Frege's puzzle that poses a significant problem for this view.
>
> (Millar 2015 p. 368)

In fact Millar's claim is even stronger: that although there is an analogous problem for object-involving theories of perception, analogous *solutions* to those often employed by Millians against Frege's puzzle are unavailable. Millar takes the general form of a case including one object (the term is used loosely—sometimes the 'object' is a person) and one subject, but where the subject ratio-

[23]Millar actually argues against the 'singular content view', according to which the content of a perceptual experience is a proposition 'partly composed of the particular physical object(s) [it] is about or of'. (p. 369) But as he says in footnote 4, his argument is directly applicable to any view where the contents of a perceptual experience include the objects perceived; this includes naïve realism as I characterise it.

nally believes there to be two, distinct objects.[24] Thus in the Trivial/Informative case above, a rational subject competent with both names might reasonably believe there to be two objects, a and b, when in fact there is only one. So a Frege puzzle for perception must be a case of a subject who (perceptually) mistakes one object for 'two' distinct objects. Millar stresses that this mistake must be in the subject's *perceptual experience*, not just in the beliefs subsequently formed.

Millar adapts an example from Perry (2001 pp. 63–64):

> Suppose there is a dog with an unusually long torso, named "Stretch", curled around a small pillar or post, and that from your vantage point a portion of Stretch's midsection is obscured by this pillar. … The result is that from your vantage point the scene in question might be perceived as a single dog positioned such that part of its midsection is occluded by the pillar, or as two stubbier dogs situated such that the head and front legs of the dog to your left and the tail and back legs of the dog to your right are occluded by the pillar.
>
> (Millar 2015 p. 374)

You observe Stretch, and have a visual experience as of seeing two distinct shorter dogs, one with its head near the other's hind legs, both obscured by the post, with the former's hind legs and the latter's head next to each other in front of the post (i.e. you have an experience determinately of the former kind in the quotation). This is consistent with the actual fact that there is only one very long dog, curled around the post. Call this experience *E*.

So we have a Frege puzzle: a perceptually well-functioning subject who mistakes one object (Stretch) for 'two' distinct objects. Frege's original puzzle was marshalled against Millianism, and this perceptual case can be brought against naïve realism in a similar way. On a view like naïve realism 'Presumably, in order to see a particular dog it's sufficient to have a clear visual experience of one side of half of that dog'. So according to naïve realism the contents of E present exactly one dog, since there is only one actual dog which you see. But this is in tension with the phenomenology of experience, which presents *two* dogs.

[24]Millar cites Salmon (1986 Chapter 7) and Thau (2002) in support of this way of putting this puzzle. This interpretation is compatible with other ways of understanding it, including the common reading where the puzzle is to explain why $a = a$ and $a = b$ afford different inferences though there is a sense in which they are equivalent. We could read 'mistaking one object for two' as doing what we should do one two objects are concerned (i.e. use distinct names $a$ and $b$, and afford the inferences this gets us) in a case where in fact only one object is concerned (so we 'should' use a single name twice, and afford the inferences that gets us). Though he doesn't discuss the difference, Millar likely adopts this interpretation because it is the more straightforward to demonstrate perceptually.

To make the comparison more explicit, we can compare E with another experience. You might also have observed Stretch, from exactly the same vantage point, in the same lighting, etc. but had a visual experience as of seeing *one* long dog, curled around the post (i.e. an experience veridical with respect to the number of dogs present). Call this $E*$. A naïve realist view will be unable to distinguish the contents of $E$ and $E*$, just as a straightforward Millian view is unable to distinguish the contents of Trivial and Informative. But clearly the contents *do* differ, so no such view is sufficient.

Millar's case is by far the clearest, but others have suggested perceptual analogues of Frege's puzzle too. G. Siegel (2022) considers an inebriated subject with double-vision: one object is presented to them twice, and it appears to them that there is a spatial relation between these 'two' objects (Siegel also cites Perry's case).

Wettstein (1986) presents two similar cases involving perception (p. 188 and p. 196), though neither is a truly *perceptual* instance of Frege's puzzle. Combining the two leads to the following more definite case. Imagine that that you are backstage at a rock concert, hoping to catch the eccentric star. The star is made up and dressed such that their left and right profiles appear very different. They are also obsessed with self-image, and their dressing room is filled with large mirrors at odd angles. Standing in the room, you have an experience E which you take to be of two separate people, facing in opposite directions. In fact, you see the one star, strangely reflected in their own mirrors. Had you seen the star beforehand, clocked the makeup, and watched the two reflections emerge in unison as the entered the room, then you would have had an experience E*, very much like E (from the same vantage point, in the same lighting conditions, etc.) but as of one object, strangely reflected. The important difference with Wettstein's own cases is that you see the 'two stars' at once—in his cases you see one then the other.

Common to all these is the same dialectic as before: two different experiences are both possible given the actual array of objects before the subject. The two differ with respect to how many objects the subject seems to see. But they do not differ in any other respect. In particular, they do not differ in how many actual objects the subject is perceptually related to. The two experiences must differ in content, but do not differ in which *objects* they relate to the subject. So, contra views like naïve realism, content cannot be the same as objects-perceived.

Sitting somewhere between cases only concerned with language or perception are applications of Frege's puzzle concerned with demonstratives. The most famous case is Perry's:

> Suppose I am viewing the harbor [sic] from downtown Oakland; the bow and stern of the aircraft carrier *Enterprise* are visible, though its middle is obscured by a large building. The name "Enterprise" is clearly visible on the bow, so when I tell a visitor, "This is the *Enterprise*," pointing towards the bow, this is readily accepted. When I say, pointing to the stern clearly several city blocks from the bow, "That is the *Enterprise*," however, she refuses to believe me.
>
> (Perry 1977 pp. 483–4)

Perry says later that 'This is, of course, simply a variation on Frege's own Morning Star example'. Evans (1982 p. 84) and Campbell (2002 p. 5) discuss similar cases.

As with failure cases, the commonality of the dialectic should by now be clear. In language, perception and thought there are cases involving two instances of some phenomenon, which present exactly the same worldly objects, but such that a rational (perhaps non-philosophical) subject could take one to present two objects, and the other only one. Many have taken this to demonstrate that the two instances differ in content. But since they do *not* differ in which objects they present, the content of such an instance cannot just be the objects presented; views like Millianism must be false. As before though, my concern here is not with this conclusion, it is with the significant similarity of the arguments advanced with respect to different phenomena.

### 2.2.1 'Paderewski' cases

Before moving on, it is worth briefly discussing 'Paderewski' cases. These originate with Kripke (1979) and are often mentioned alongside Frege's puzzle. Here is one of Kripke's original cases:

> Peter … may learn the name 'Paderewski' with an identification of the person named as a famous pianist. Naturally, having learned this, Peter will assent to "Paderewski had musical talent," and we can infer — using 'Paderewski,' as we usually do, to name the Polish musician and statesman:
>
> (8) Peter believes that Paderewski had musical talent.
>
> … Later, in a different circle, Peter learns of someone called 'Paderewski' who was a Polish nationalist leader and Prime Minister. Peter is skeptical of the musical abilities of politicians. He concludes that probably two people, approximate contemporaries no doubt, were both named

> ‘Paderewski.’ Using ‘Paderewski’ as a name for the statesman, Peter assents to, “Paderewski had no musical talent.” Should we infer, by the disquotational principle,
>
> (9) Peter believes that Paderewski had no musical talent
>
> or should we not?
>
> (Kripke 1979 p. 265)

The puzzle is that although Peter is not irrational, and his beliefs are entirely banal, there does not appear to be a sensible way to characterise them. We cannot say he simply believes Paderewski has musical talent, because this privileges the first claim (8) over the second claim (9), though they are arrived at by identical methods. For converse reasons we cannot say he simply believes Paderewski has no musical talent. We cannot say he believes neither, since he clearly does hold beliefs about Paderewski and his musical abilities. But we cannot say he believes *both* as this commits Peter to irrationally believing that Paderewski has and does not have musical talent, which is absurd. These options are logically exhaustive, but the situation is very straightforward. The puzzle is to provide a sensible characterisation without absurdity. Similar puzzles are considered by Quine (1956 p. 179) and Salmon (1986 Chapters 7-9).

As Kripke’s paper makes clear, the puzzle is about *belief*, or perhaps about belief attributions. Many interpreters of Frege’s puzzle have been very interested in what it tells us about belief and belief attribution, and how these relate to the topics of this thesis. (e.g. Katz 1994; 2001; Jeshion 2000; Soames 2002) This thesis isn’t about belief, so I won’t engage with the puzzle as Kripke presents it here.

Cases of this kind have also been put to another use: to demonstrate that instances of Informative do not require two different signs, because there are non-trivial claims of the form ⌜a is identical to a⌝. (e.g. Sainsbury 2005 pp. 67, 103ff; 2003b) The idea is that a person like Peter can ‘without obvious irrationality, believe something he could express as ‘Paderewski is not Paderewski’ (Sainsbury 2003b p. 134) or equivalently, that he could rationally deny something he would express as ‘Paderewski is Paderewski’, because for Peter these claims use two different (though qualitatively identical) names.

I am rather sceptical of the philosophical importance of this, as I am sceptical of how often it actually *happens*. Presentations of apparent examples often include an appeal to context, or additional qualifiers like ‘the pianist’ or ‘the statesman’ to distinguish the two uses of names, and so do not really establish that exactly the *same* sign can appear on both sides and the identity still be informative. Nevertheless, it is clearly possible that, on considering ‘both’ men for

some reason, Peter could say to himself, 'ah but I must remember that Paderewski is not Paderewski' and express something sensible.

This is an uncommon interpretation of a recent, and minor, part of the debate, so it does not play a role in my arguments for Unity or the nature of the bearers of content. However, it is a useful test of the eventual success of my theory that I can explain how such cases are possible. I provide an account of this in §§4.3.3.3 and 5.4.4.1.

## 2.3 Alternative-satisfaction cases

Failure cases and Frege's puzzle have been presented as powerful objections to object-dependent theories of content. A common response is to abandon object-dependence entirely, in favour of 'internal' content. Theories have been developed on which the content of a name, thought or experience is not (and does not depend upon) a worldly object. Instead, the content is some description, property or cluster of properties, which is *independent* of any object. The content is supposed to specify or individuate the object as that which has these properties (all of them, some of them, a weighted most of them, etc. on different theories). If the content is in fact satisfied by a certain thing, then that thing is the worldly object to which the thing which bears the content relates. Notice that such views include two distinct claims: one about the nature of content, and the other about the relation between content and worldly objects or referents.

Views like this account easily for failure cases: satisfiable content can exist without in fact being satisfied. The content of a name, thought or experience is just a property or description, and it is *this* which gives it its significance, not the object (if any) which satisfies that property or description. Whether there even *is* such an object may be entirely irrelevant to the theory.

Frege's puzzle dissolves in a similar way: two different names (thoughts, experiences, etc.) might have different properties or descriptions as their content, though these properties or descriptions are satisfied by the same thing. So the content of 'Hesperus' might be the property of being the brightest heavenly body which rises in the evening, and the content of 'Phosphorous' that of being the brightest heavenly body which rises in the morning. It so happens that these properties are both satisfied by the same thing, but this is irrelevant, and one can still know the content of each name without knowing this further fact. For such a person Trivial will be obvious, but Informative won't be. This is (very roughly) how Frege himself solves the puzzle.

All very well. But there seem to be cases of names, singular thoughts and perceptual experiences which fulfil the following criteria: 1) it is fairly clear what

the satisfaction conditions of their content would be were a satisfiable-content theory correct, 2) the conditions are satisfied by some object O, 3) it is clear that O is not the referent (either because it is clear what the unique referent is, or because it is clear that there is no referent at all). The theory is thus committed to this other object O being a/the referent of the name, thought or experience, when in fact it is not, so the case is a counterexample to the theory. I call these *alternative-satisfaction* cases.

Some concrete examples will make this more clear. As before, we will see that this dialectic occurs across philosophical work on language, singular thought and perception.

The most famous example of this move is Kripke's (1980) argument against what he calls the 'Frege-Russell view' (a crude version of Russell's descriptivism or Frege's senses). According to the Frege-Russell view as Kripke understands it:

> 1. To every name or designating expression 'X', there corresponds a cluster of properties, namely the family of those properties $\varphi$ such that A [the speaker of the name] believes '$\varphi$X'.
> 2. One of the properties, or some conjointly, are believed by A to pick out some individual uniquely.
> 3. If most, or a weighted most, of the $\varphi$'s are satisfied by one unique object y, then y is the referent of 'X'.
>
> (Kripke 1980 p. 71)

Though it does not appear in this list, Kripke also ascribes something like the converse of 3 to the Frege-Russel view. He says 'The picture associated with the theory is that only by giving some unique properties can you know who someone is and thus know what the reference of your name is', i.e. that satisfaction is necessary for reference, not just sufficient for it as in 3. This remark also commits the theory to:

4. If there is a unique referent of 'X'—call it y—then y is the unique satisfier of most, or a weighted most of the $\varphi$s.

4 corresponds pretty well to the second claim of the satisfiable-content theories discussed above. An alternative-satisfaction case is a counterexample to 4. As Kripke points out, there are names which are clearly associated with certain properties by their speaker, and which clearly do have a determinate referent, but where the referent is not the unique satisfier of most (or a weighted most) of the properties. Of particular interest in this section are cases where the referent is not the *unique* satisfier because something else satisfies the properties (perhaps as well as the actual referent satisfying them).

Kripke's own example is 'Cicero': 'I think you do know who Cicero is if you just can answer that he's a famous Roman orator'. The idea is that you (the man in the street, the non classically-educated, etc.) associate nothing more with 'Cicero' than the property of being a famous Roman orator. Though this property is satisfied by Cicero, it is not satisfied by him *uniquely*—there were many other famous Roman orators (Pompey, Cato, Horace). Yet you still manage to refer to Cicero, and nobody else, when using the name 'Cicero'.

I find Kripke's 'Feynman' case a clearer example. He brings this up as a counterexample to 2, but it works against 4 equally well:

> Consider Richard Feynman, to whom many of us are able to refer. He is a leading contemporary theoretical physicist. Everyone here (I'm sure!) can state the contents of one of Feynman's theories so as to differentiate him from Gell-Mann. However, the man in the street, not possessing these abilities, may still use the name 'Feynman'. When asked he will say: well he's a physicist or something. He may not think that this picks out anyone uniquely. I still think he uses the name 'Feynman' as a name for Feynman [and not for Gell-Mann].
>
> (1980 p. 81)

Again, the man in the street associates a certain property (being a famous physicist) with a name ('Feynman'). They associate no other properties with the name. Though Feynman *was* a famous physicist this does not individuate him uniquely; there have been other famous physicists (Einstein, Oppenheimer, Gell-Mann).

So 'Cicero' and 'Feynman' provide counterexamples to 4, because they *do* have a unique referent but the properties associated with them are not uniquely satisfied by that referent. The cases show that name-reference does not require unique property-satisfaction—the second claim of the object-independent theories of names is false.

The label *alternative-satisfaction* should now make sense. There is an 'alternative' particular—not the referent—which in fact 'satisfies' the properties supposed (by the internal-content view) to uniquely individuate the referent. So the internal-content theory is wrong.

The main purpose of this section is to demonstrate that the pattern of raising alternative-satisfaction cases against object-independent theories occurs with exactly the same role and force, in debates about the content of names, singular thoughts, and perception. Kripke's cases are so well known that philosophers of language interested in such problems simply cite them (though see Ziff 1977

p. 22 on 'Andasill' and Strawson 1964 p. 20 on massive reduplication). Let us turn to cognitive cases.

Cognitive alternative-satisfaction cases are singular thoughts which—were a satisfiable-content theory true—would be associated with properties satisfied by one particular thing, but which are clearly not about that particular thing.

Zarepour (2018) provides a clear example. He begins with a brief discussion of Crane's 'psychologistic' account of intentionality.[25] The important point for Zarepour is a contrast between reference and *aboutness*:

> 'Aboutness' is the mere representation of something in words or thoughts, whether or not it exists. So although my word 'Pegasus' does not refer to the mythological winged horse Pegasus—'Pegasus' is, after all, commonly called a 'non-referring term'—I can talk or think about Pegasus.
>
> (Crane 2013 p. 9; quoted in Zarepour 2018 p. 4)

Zarepour provides the following counterexample:

> Suppose that through the window of my flat, in a specific situation and at a particular time of day, I see John walking in the street. Based on this perceptual experience, some thoughts about the object of my experience appear in my mind. But, unfortunately, at the end of that day I forget anything happened and any thought I had. Suppose moreover that the day after I have exactly the same experiences and thoughts, this time of Jones, John's twin who looks like him in all respects. At the same time of day and in the same situation, through the same window of my flat, I see Jones walking in the same area. Based on this perceptual experience, similar thoughts about the object of my experience appear in my mind. Finally, like the first day, at the end of the second day I forget everything happened and any thought I had on that day.
>
> (Zarepour 2018 p. 5)

The argument against psychologism then goes like this. On each day, I have a singular thought. According to psychologism, the thoughts are not directly associated with John or Jones, but have an internal representational structure independent of them (presumably, one which picks out John/Jones as the satisfier of some properties). Since the internal and external conditions under which these two thoughts occur are subjectively indistinguishable, the representational

[25] Zarepour cites Crane (2001; 2013) and reprints in a later collection of essays.

structure generated in each case must be the same. Since the aboutness of each thought is a result of its representational structure, psychologism predicts that the two thoughts should be about the same thing. But clearly they are *not* about the same thing—my thought on the first day is clearly about John (and not Jones), and that on the second day about Jones (and not John).

The thought on the second day, about Jones, constitutes an alternative-satisfaction case. This thought has a certain representational structure associated with it—perhaps something like a representation of a man, walking at a certain speed, in a leather jacket and baseball hat, etc. This representation (and nothing but this representation) determine what the thought is about. Though Jones satisfies the representation (so the thought is about him), the representation does not represent him uniquely—it is equally applicable to John. But the thought on the second day is about Jones and *not* about John.

So the thought on the second day provides a counterexample to psychologism, because it is about only one thing, but the representational structure associated with it is not uniquely satisfied by that thing. The case shows that the aboutness of singular thoughts does not require unique property-satisfaction (contra object-independent theories of content, like Crane's psychologism).

Notice the clear parallel with 'Cicero'. In fact, 'Cicero' refers to exactly one person—Cicero. But if we take the Frege-Russell theory at its word, it would seem that 'Cicero' might reasonably refer to any Roman orator, and there is nothing to decide which. Similarly, the thought on the second day is in fact about Jones, but by the lights of psychologism it might as well be about John, and there is nothing to decide which.

Evans considers a very similar case, though in service of an orthogonal point. He wants to explain *Russell's Principle*, the claim that for a subject to make a judgement about some thing, they must have discriminating knowledge of that thing ('a capacity to distinguish the object of [their] judgement from all other things'). (1982 p. 89) In fact, Evans applies Russell's principle beyond judgement, to discussions of 'thoughts' (not merely judgements, e.g. p. 91), demonstrative identification and understanding of utterances (p. 92) and the meanings of names.[26] Discussion of the principle pervades the book and it is clear that if Evans held any views on the questions of content I am concerned with, Russell's principle would be relevant. In attempting to precisify his initial characterisation, Evans considers an alternative satisfaction case:

[26]On page 92 Evans quotes a passage from Dummett (1973) on the meaning of names, as part of his continuing discussion of Russell's principle. Evans himself doesn't explicitly distinguish discussions of names and thoughts here. On page 98 he discusses the name 'Caesar' as if Russell's principle could be applied to it (either directly or by being applied to some related entity like a proposition, derived from a sentence or utterance containing the name).

> our subject briefly sees one ball rotating by itself on one day, and the other on a later day. And let us further suppose that the subject retains no memory of the first episode, because of a localized amnesia produced by a blow to the head. Suppose, finally, that many years later our subject reminisces about 'that shiny ball' he saw many years earlier. If asked which ball he is thinking about, our subject cannot produce any facts which would discriminate between the two. (I assume, as I think I am entitled to, that he would not think of distinguishing the ball he is thinking of as the one from which his current memory derives.) [even though the subject believes he can discriminate the ball by its properties, and in fact could not] it would certainly be quite natural, in view of the facts, to say that he was *thinking of* the second ball, or that he had the second ball *in mind*, or, if he spoke, that he *meant* the second ball.
>
> (Evans 1982 p. 90, original italics)

Evans' case works like Zarepour's. There are two objects which cannot be subjectively distinguished. The subject observes one, then the other, and forgets all about the first observation. Then the subject thinks about what seems to them to be the (only) object with the relevant characteristics. The intuition we are invited to share is that they really think about the later object—which they remember—not the early one. But this cannot be in virtue of any object-independent representational capacity, because any such capacity would individuate both objects, since they share all the features by which it might do so. So knowing which object one is judging about (thinking about, naming, discussing, etc.)[27] cannot depend merely on object-independent representational capacities.

The important points for my discussion are that Evans considers a case of this kind, clearly believing it to be philosophically telling, and in particular, believes it to tell *against* a conception of Russell's principle in terms object-independent representation.

Finally, let us turn to perceptual alternative-satisfaction. A perceptual alternative-satisfaction case is a perceptual experience which—were a satisfiable-content theory true—would be veridically of a certain thing, but which is clearly *not* veridically of that thing.

There is a rich tradition of such cases. Grice's (1961) clock is probably the best known example. Suppose X is a subject with normal, functional, visual

[27]The only feature they *don't* share by which a representation might individuate one is being-the-cause-of-that-representation, which Evans assumes neither includes. Appeals to such self-referential properties also appear in other literature (e.g. rigidified definite descriptivism and Searle's existentially quantified theory of intentionality).

apparatus. There is a treatment 'by which an expert could make it look to X as if there is a clock on the shelf on occasions when the shelf was empty'. It looks to a subject X as if there is a clock on the shelf, and there really *is* a clock on the shelf (which really does look the way X seems to see a clock looking). But, unknown to X they are under the treatment.[28] We are invited to share the intuition that if X's impression of there being a clock remains unchanged even when the clock is (re)moved, then X never really *saw* the clock at all. A less widely-cited but perhaps clearer example follows immediately:

> it might be that it looked to me as if there were a certain sort of pillar in a certain direction at a certain distance, and there might actually be such a pillar [P1] in that place; but if, unknown to me, there were a mirror interposed between myself and the pillar, which reflected a numerically different though similar pillar [P2], it would certainly be incorrect to say that I saw the first pillar, and correct to say that I saw the second;

Both cases work the same way. There is a fact of the matter about what the agent sees; it is unclear in the first case, though it is not the clock, and it is clearly P2 in the second. The agent has a phenomenal experience which—with respect only to the distribution of properties and objects apparently seen—is satisfied by some object which they do not see. The point for Grice is to demonstrate that a causal connection between the object seen and the subject's perceptual experience is necessary for them to really *perceive*. In particular, that mere 'fit' between the properties and objects which the subject seems to see and what is actually there is *not* sufficient in the absence of such a causal connection.[29]

Comparison with 'Cicero' should be clear. By the lights of the Frege-Russell theory, what a name refers to is just a matter of what properties it is associated with and whether they are satisfied. But there are cases where a name is associated with properties, and those properties are satisfied by a certain thing, and yet it is clear that the name does *not* refer to that thing. Similarly here, if what the subject sees were just a matter of what properties and objects they seem to perceive and whether this seeming is satisfied, then the subject would see the clock. But the subject clearly does not see the clock, even though it instantiates all the right properties. So, what a subject sees cannot be merely a matter of what properties and objects they seem to see, and whether these are satisfied (compare: 'Cicero' clearly does not refer to Horace though he too was a Roman

[28]Grice doesn't make this explicit, but it is pretty clear that whether there really *is* a clock does not alter the effect of the treatment.

[29]The latter point is important for my discussion, the former is not.

orator, and the thought on the second day is not about John, even though he too was wearing a leather jacket and a baseball hat).

To return briefly to the section heading, Grice's clock is an 'alternative satisfaction' case because there is an object—the actual clock—which satisfies a purely qualitative description of what X seems to see, but which is *not* what X actually sees; it is an *alternative*. It is irrelevant to my characterisation of this as an alternative-satisfaction case that in fact X does not see any clock at all (indeed it is unclear from Grice's description what X does see).

The pillar case makes this more clear. Grice only actually sees P2, but by the lights of the theory implicitly under attack he sees P1, because P1 satisfies a purely qualitative description of what he seems to see. (The theory implicitly under attack might entail that Grice does *not* see P1, but this is irrelevant to my characterisation of this example as an alternative-satisfaction case.)

There are many similar cases. Evans (1973) and Altham (1973 a response to Evans) both discuss the fact that a visual experience's fitting some object in the world is not sufficient for perception—the object must cause the experience. Altham is particularly explicit:

> It is perfectly evident that the fact that there is an excellent match between a person's visual impression and a material object is not a sufficient condition of his perceiving that object, and this truth can be made an important part of a strong case for claiming that for the object to be seen, it must in some way be causally operative in producing a visual impression.
>
> (Altham 1973 p. 210)

Burge says similarly 'It is in principle possible for identical visual fields to be associated with different referents'. (1979 p. 402)

Searle's (1983) discussion of Bob and Sally runs on similar lines, though bears further discussion. Searle thinks that the content of a visual experience is exhausted by its conditions of satisfaction, and that upon seeing an object of a certain type, these will be that there exists such an object of that type where there appears to be, and that the fact that there is such an object there is causing the experience. So on seeing a yellow station wagon, I have a perceptual experience with the content *I have a visual experience (that there is a yellow station wagon and that there is a yellow station wagon there is causing this visual experience)*. (ibid. pp. 48-9, see also pp. 61-2) As in Crane's psychologism and the purely qualitative views Grice discusses, no actual station wagon, nor any reference to one, enters the description of the content of a perceptual experience. But Searle recognises a serious problem for his view:

> Suppose that in a distant galaxy there is the twin of our earth, type identical with our earth down to the last microparticle. Suppose that on our earth Bill Jones sees his wife, Sally, getting out of their yellow station wagon and on twin earth, twin Jones sees twin Sally getting out of their twin station wagon. Now what is it about the *content* of Jones's visual experience that makes the presence of Sally rather than twin Sally part of the conditions of satisfaction of his visual experience, and what makes the presence of twin Sally part of the conditions of satisfaction of twin Jones's experience? By hypothesis both experiences are qualitatively identical, yet it is part of the conditions of satisfaction of each agent's experience that he is not seeing just *any* woman with such and such visual characteristics, but that he is seeing his very own wife, Sally, or twin Sally as the case might be. … The question, to repeat, is: How does the particularity get into the Intentional content?
>
> (Searle 1983 pp. 62–3)

The structure is familiar: the real Jones has an experience whose content *actually* concerns the real Sally. But the conditions of satisfaction would be satisfied equally well by Twin Sally too—what makes it the case that only the real Sally is relevant?

A brief note on causation: Searle's inclusion of a causal constraint in the conditions of satisfaction might cloud the issue. In the two pillars case above, the 'satisfaction conditions' of the experience are (according to Grice) just that there is a pillar in a certain place. The problem is that there is a pillar (P1) other than the one Grice really sees (P2), which satisfies these conditions. Grice solves the problem by claiming that causation is necessary for seeing (perhaps as well as condition-satisfaction). On Searle's account, causation is *one of the conditions of satisfaction*—a Searlean description of the content of Grice's experience would be *I have a visual experience (that there is a pillar and that there is a pillar there is causing this visual experience)*. One might think that Searle is immune to alternative-satisfaction concerns, since it is specified in the satisfaction conditions that the relevant thing must be causing the visual experience, and in his case it is only the real Sally who is *really* causing the experience. Not so. It is true that Jones sees Sally just if Sally causes his experience (and Twin Sally just if she causes his experience), but Searle's concern is why *this* is so. As Searle puts it 'what is it about his visual experience that requires Sally and not somebody type identical with Sally [to be causing it]?'. The question might be put better this way: had Jones had an experience with exactly the same satisfaction conditions, but *Twin*

*Sally* had been in front of him, we should want to say that the conditions were not met, but in virtue of what could we say this? This is a substantive concern. The most significant difference between Grice's case and Searle's is that Grice's includes an *actual* particular which alternatively satisfies the relevant conditions, whereas Searle's contains only a *possible* one. In important respects though, the two are structurally alike.

Searle is even aware of the ubiquity of this problem across domains: 'The problem of particularity crops up in a variety of places in the philosophy of mind and philosophy of language'. He also notes (p. 62) that his case is a variant of Putnam's (1973) Twin Earth, which in turn is often compared with Kripke's cases discussed above (I do not discuss Putnam's case because it concerns kinds rather than particulars).

Soteriou (2000) presents a similar case against a simple form of intentionalism, where:

> The intentionalist claims that a visual experience is a mental state with an intentional content that represents the world as being a certain way. The content of an experience can be correct or incorrect – veridical or non-veridical.
>
> (Soteriou 2000 p. 173)

Soteriou argues that a purely qualitative intentionalism is insufficient, with the following case. There is a red ball B1 to a subject's left, a similar red ball B2 in front of her, and a third similar red ball B3 to her right. She is wearing 'displacing glasses' which make things appear further right than they are—so much further right that B1 affects her phenomenology as if there a red ball in front of her (where B2 really is), B2 as if there is a ball to her right (where B3 really is) and B3 does not affect her. For some reason (Soteriou doesn't specify) 'it does not seem to the subject as if there is no red object to her left'. So with respect to the various balls, the subject's experience is entirely veridical: it seems to her that there is a red ball in front of her (where there is; B2) and a red ball to her right (where there is; B3). But intuitively the subject is *not* seeing the red ball (actually) to her right, and if she sees the balls to her left and front at all, she sees them in the wrong places—her experience isn't actually veridical. But this latter claim is not consistent with a entirely qualitative view of the content of perception. With respect to B2 this is an alternative satisfaction case: there is a fact of the matter about which object the subject sees (B2).[30] The subject has a phenomenal

[30]Similar comments apply to B1, perceived as if where B2 is. The fact that we must case is best explained in terms of one object at a time is telling, and will be a major theme of the next chapter.

experience which the qualitative intentionalist says is verified under certain conditions (if there is red ball to her right), and those conditions *are* fulfilled, but by a different object B3. Soteriou concludes that intentionalists 'should hold that the truth-evaluable contents of such experiences are object-involving'.

In a footnote, Soteriou credits his understanding of such examples to Martin, who has discussed a variety of similar cases (e.g. in 2002a). The most applicable here is the 'veridical hallucination' case of an orange (p. 9): a subject hallucinates an orange on the table before them, and (entirely unrelated to the hallucination) there *is* an orange on the table before them, which does not perceptually affect them. With respect merely to the 'general' content of their experience they 'see' the real orange, though clearly they do *not* really see it. Martin also considers variations concerning properties (the pink/white candle), and objective continuity (the paper crosses) on the following pages, and also cites Grice's paper.

Johnston (2004) considers a number of similar cases. A patient lies on an operating table and their brain is artificially stimulated so that they seem to perceive lights in a ceiling above them. In fact, there *are* lights in exactly the places in the ceiling where the patient seems to perceive lights, but intuitively, they do not see the real lights (p. 122). Johnston concludes the content of the experience cannot be simply that there are lights in certain places[31] of certain brightness, etc. because these conditions are all fulfilled by the real lights in the real ceiling; the real lights are the alternative satisfiers of the experiences' qualitative content. Johnston's parallel case of the aunt calling on Friday (p. 132) is the only alternative-satisfaction case I have found developed in auditory terms.

Cases of a similar structure, with no noteworthy differences, are also discussed by Pears (1976) and Campbell (2002 p. 125); by Crane (2005 p. 258) and Tye (2007 p. 596) who both note the relation with Grice; by Burge (2009 p. 293) in explicit comparison with Kripke; by Bach (2010 p. 40) who notes the parallel with singular thought; and by Schellenberg (2010 p. 22; 2016 p. 31, p. 35).

This concludes our discussion of alternatives satisfaction cases. Its purpose was to demonstrate that—as with failures cases and Frege's puzzle—alternative satisfaction cases are deployed against object-independent theories, with exactly the same role and force, in debates about the content of names, singular thoughts, and perception. Writers in all three domains have brought to bear cases of an apparent bearer of content, whose content's satisfaction conditions (were a satisfiable-content theory true) are clear, but which do not refer to the/a thing which satisfies these conditions.

[31]Or with respect to any one apparently seen light, that there is a light there.

## 2.4 Non-satisfaction cases

The fourth and final kind of case is *non-satisfaction*. Like alternative satisfaction cases, these are deployed as counterexamples to object-independent theories of content. The two kinds of case are often presented together (or simply conflated). In an alternative-satisfaction case, it is clear what the conditions of satisfaction are or should be, and it is clear what the referent is, and the problem is that there is some object which fulfils the conditions but isn't the referent. A *non-satisfaction* case is a name, singular thought or perceptual experience which fulfils the following criteria: 1) it is fairly clear what its content's satisfaction conditions would be were a satisfiable-content theory correct, 2) it is fairly clear what the referent is and 3) the referent does not satisfy the conditions.

Importantly, this last condition is entirely independent of whether anything *else* satisfies the conditions, so whether a scenario is an alternative satisfaction case is independent of whether it is an non-satisfaction case are independent.

Again, Kripke provides canonical examples, though they are slightly less clear, since all non-satisfaction cases which Kripke considers are also alternative-satisfaction cases. Kripke says that all a layman could tell you about Columbus is that he was the first European to land in the Western Hemisphere, or the first man to realise that the earth is round. Neither is true, but users of the name still manage to talk about *Columbus* (and only Columbus) by using the name.[32] So 'Columbus' provides a counterexample to the satisfiable-content theories: according to the theories, lay uses of 'Columbus' should refer to whatever satisfies the properties associated with them (i.e. the first European in America), but in fact they refer to Columbus, who does *not* satisfy these conditions; hence the name 'non-satisfaction' case. Kripke's Gödel/Schmidt and Peano/Dedekind cases work in exactly the same way (pp. 83–85). All three also constitute alternative-satisfaction cases because the relevant conditions *are* fulfilled by other people.

It would be easy to construct a purely non-satisfaction case along similar lines. Alice attends a popular lecture on great problems in the history of mathematics. She hears about Fermat and his theorem, about Andrew Wiles who proved the theorem, and finally about the Collatz Conjecture which remains unproven. But things get muddled in Alice's memory, and she leaves associating the name 'Andrew Wiles' with the property *prover of the Collatz Conjecture*. If Alice tells a mathematically-knowledgeable friend that she has learned about 'Andrew Wiles, who proved the Collatz Conjecture', the friend will explain that she is wrong, and Wiles actually proved Fermat's last theorem. But for the friend to be

[32]The first Europeans to land on mainland present-day America were Leif Erikson and his party, viking explorers taken to Greenland by Erikson's father. It is debatable who first realised the earth was round, but they certainly lived before Columbus.

justified in this correction, Alice must really *be* referring to Wiles, in spite of the fact nobody at all has proved the Collatz Conjecture. So, contra the satisfiable-content theories, it can't be that the content of 'Andrew Wiles' is just the property *prover of the Collatz Conjecture*.

Devitt (1981 pp. 17–18) cites Wittgenstein's (2009 §79) discussion of Moses as a non-satisfaction case. Suppose, as is likely, that nobody did all or even most of the things the Bible attributes to Moses (and suppose that none of these things is privileged in our modern conceptions of Moses). Nevertheless, if there really was an historical person, whose name has come down to us as 'Moses', then our utterances of that name refer to him, even though most of what we believe about him is false. Devitt also considers the case of 'Jonah', to whom we may refer although he did not survive inside a whale, and describes how this mismatch between fact and belief could have originally come about.

The 'Columbus', 'Moses', 'Jonah' and similar cases can all also be cast in terms of singular thoughts about Columbus, Moses, Jonah etc., by those who hold only the above beliefs, so Kripke's and Wittgenstein's points apply to cognition too.

Evans' careful (1973) discussion of Kripke's attack on descriptivism is also instructive. Evans points out that Kripke's arguments go through only if:

> First: that in order to be saying something by uttering an expression one must utter the sentence with certain intentions; this is felt to require, in the case of sentences containing names, that one be aiming at something with one's use of the name. Secondly—and this is where the underpinning from a certain Philosophy of Mind becomes apparent—to have an intention or belief concerning some item (which one is not in a position to demonstratively identify) one must be in possession of a description uniquely true of it.

Presumably a singular thought about an object is or involves such 'aiming' at that object, but without the associated name-utterance. If Evans is right then, any of Kripke's examples can be taken just as well as cases of singular thought, by just considering the mental processes which would accompany the speaker's saying the name.

There are also perceptual non-satisfaction cases. Campbell (2002) considers several. The clearest is his discussion of the success or failure of 'binding parameters' and 'bundling principles'. Campbell thinks there are 'binding parameters'—properties like location or colour—which organise how attention is distributed by visual mechanisms and how instances of different characteristics are treated as belonging to the same object. But importantly, these characteristics are only *used* by the subject in collecting information together, they can

'discharge their work without providing descriptive conditions which an object must meet in order to be the object to which you are referring'. (p. 38) Thus, if location is a binding parameter, then:

> you could be using the apparent location of an object perfectly well in sorting together all the visual information that relates to that object, whether or not the object is actually where it seems to be, rather than being reflected in a mirror or seen through a prism.
>
> (Campbell 2002 p. 38)

This is clearly a non-satisfaction case: the object does not in fact instantiate the relevant properties (being in a certain location) yet one still sees the object. Campbell is very aware of the analogy with names. He writes:

> The analogy I have in mind is with the way in which gossip, or, more generally, information about individuals circulates in a community. All this information has to be sorted into bundles, so that each bundle contains information relating to just one individual; otherwise we would never know whether we were talking about the same or different people. The basic method we have for doing this is the use of proper names: we try to sort so that sameness of name is an indication that it is the same individual that is in question. That is not quite enough, though, since different people may have the same name. So we may add a tag. For example, if there are many different people named Williams, we might add tags like 'the milkman' or 'the mathematician'. Then the use of the name plus a tag gives us what we might call a bundling principle: a way of sorting together information as all relating to a single individual.
>
> …
>
> bundling principles can do the work they do whether or not they are actually true of the object. It is perfectly coherent to suppose that Williams the mathematician might turn out not to be a mathematician at all, but a criminal who pretends to be engaged in mathematics. The tag, 'the mathematician', could still perfectly well play its role
>
> (Campbell 2002 pp. 38–9)

This too is clearly a non-satisfaction case: however names-with-tags like 'Williams the mathematician' really function, it is clear that they do not refer to the bearer by satisfaction since—*ex hypothesi*—their bearers do not always satisfy the relevant properties.

Campbell also cites Shoemaker (1968 cited on p. 91). Shoemaker points out that, although one might use apparent features of an object to demonstrate *it* (rather than some object available for demonstration), whether the object actually *has* these features is irrelevant to whether the demonstration succeeds. Campbell puts it 'judgements about the location of a demonstrated object may be mistaken, but they are nonetheless, in Sydney Shoemaker's phrase, immune to error through misidentification'.

In Shoemaker's case, I wrongly believe that there is a red silk necktie at a certain place on a shelf just behind me. I put my hand out to that place on the shelf, expecting to find a red silk tie, feel a silk tie and say 'this one is red'. In fact the tie I touch is not red. There is a sense in which I have 'misidentified' the object I refer to. But importantly I *do* refer to one particular object—the object in the location where my hand actually is (which happens to be a silk tie, though not red). This is perhaps a *partial* non-satisfaction case: I believe there is a red, silk tie at a certain location, so the object I am 'aiming' at can be specified by four properties: being a tie, being silken, being at that location, and being red. The object I actually refer to has three of those properties (it is a silk tie at that location), but lacks the last, and there is nothing which has all four. In spite of this I still manage to refer to it.

Burge (2009 p. 293) also briefly discusses non-satisfaction cases and notes similarities with what he calls 'the Kripke-Donnellan points about reference'.

By the now the plot should be familiar. What is important for my discussion of non-satisfaction cases is that the *same kinds of cases* are brought up in philosophy of language, of cognition, and of perception, and with the same dialectical role.

## 2.5 Justifying Unity

In this section, I describe why I think Unity is true. Recall the definition:

**Unity** *The nature of the significance of names, perceptual experiences and singular thoughts is the same.*

It is widely accepted that the significance of names, perceptual experiences and singular thoughts has something to with their relation to worldly objects. Most authors (myself included) demonstrate this acceptance by comfortably using the word 'reference' for the relation between a name, experience or thought and the object it is related to. It is also widely accepted that it is conceivable that names, experiences and thoughts be associated with non-objective 'general' contents, such as descriptions or properties. Much of the debate over the nature

of content concerns the relation (if any) between these two kinds of relation or association that names, experiences and thoughts might bear.

In §§2.1 to 2.4, I considered four kinds of case. Failure cases are supposed to show (against object-dependent theorists of content) that reference is not necessary for content, since there are apparent cases of content without reference. Frege's puzzle is supposed to show (against the same theorists) that reference is not sufficient for content, by showing that coincidence of reference does not entail coincidence of content. Alternative satisfaction cases are supposed to show (against object-independent theorists of content) that the satisfaction of general properties is not sufficient for reference. Finally, non-satisfaction cases are supposed to show (against the same theorists) that the satisfaction of general properties is not necessary for reference.

My goal in §§2.1 to 2.4 was to establish two distinct but related points. First, each of the four puzzles arises throughout the literature on the meaning of names, the content of perception, and the content of singular thoughts. Second, in each domain, the dialectical force of each puzzle is the same—they are *used* in the same way; Failure cases and Frege's puzzle against object-dependent theories, alternative- and non-satisfaction cases against object-independent ones. This is why it was possible for me to summarise the puzzles above without mentioning any particular phenomenon. Failure cases just show that reference is not necessary for content, and they show this regardless of what kind of thing is supposed to refer or have content.

I claim that this is all we need to have good reason to believe Unity. In the next section I describe an epistemological argument for the claim which can be constructed out of these materials. Though compelling, it is not the argument I ultimately stake my claim on. In the section after I describe a more powerful methodological argument.

### 2.5.1 An epistemological argument

From this alone follows an epistemological claim that Unity is quite likely. Why else would the same puzzles show up with the same (significant!) force in these otherwise disparate areas? Consider an analogy. I observe a small sports car which won't start. Its owner opens a hatch at the back, pours in a liquid, and then the car *will* start. Then I see a large cargo truck which won't start. Its owner adds a liquid from a much larger can to a different part of the truck, and afterwards the truck will start. Finally I observe a boat which won't go. The owner adds liquid to hatch, similar to the car's, and it sails away. Other things equal, it is reasonable to assume that all three vehicles had the sort of problem,

namely deficiency of whatever makes them move (fuel). Similarly for content: other things equal, the enormous similarities of the content-related problems above should lead us to believe Unity.

Of course, the inference is defeasible. It is *reasonable* to assume that all the vehicles needed the same thing (and that thing was fuel), but it could be otherwise. Perhaps the truck has a checking device which prevents ignition unless it has sufficient brake fluid, and the car has a similar one for screenwash (and the boat really did need fuel). And trucks, cars and boats often use quite different *kinds* of fuel. Nevertheless, we should think it fairly likely that Unity is true.

I mention this because I believe that a more or less tacit version of this argument is what has actually *caused* many philosophers to believe Unity. It was certainly the initial motivation for this project. It has some force, but it is not the argument I will rely on.

### 2.5.2 The methodological argument

Now for my argument. First, an argument like that in the previous section will demonstrate that there must be *something* in common between language, thought and perception. The four puzzles are simply enormously pervasive in the literature. There is no better way to demonstrate this than to summarise the literature, at length, as I have above. Discussion of these puzzles is not limited to any specific part or era of the literature, nor to those with certain views or background assumptions. Everyone interested in the semantics of names must at some point confront failure cases. This has been true since at least Frege's time, and it continues to be true; likewise for the other puzzles and domains.

Moreover, many of the authors cited above, and throughout the rest of this thesis, take themselves to be in dialogue with one another, both within and across the boundaries between philosophy of language, perception and cognition. Evans (1982), on language and thought, responds to Frege and Russell's earlier work, and Martin (2002a) on perception responds to Searle. But Martin also takes himself to be responding to Evans and Frege, writing on thought and language. Evans himself points out a similar tendency:

> It should be mentioned, as something of a historical accident, that a good deal of the work carried out [following Russell's concern with names and descriptions] has not been immediately concerned with distinguishing different kinds of *singular terms*. Rather, the concern has been to draw a distinction between two different kinds of beliefs about objects: *beliefs* on the one hand that are reportable *de re*, and

> beliefs on the other that are reportable only *de dicto*. Nevertheless, it was implicit in much of the discussion that the condition of being *'en rapport with'* an object which would sustain beliefs about that object reportable *de re* would also provide the condition for making and understanding references in the appropriately 'direct' style.
>
> (Evans 1982 p. 66, original italics)

All of this demonstrates that there must be *something* which language, thought and perception have in common (as investigated by analytic philosophy). It remains to demonstrate that this commonality is the nature of their contents.

Not only do the same puzzles crop up across the literature, but as I have described already, the dialectic in which they play a role is the same. Recall my rough distinction between internal and external content, and the privileged notion of content *simpliciter*. Across the board, failure cases and Frege's puzzle are put up against claims that content is just external content, and alternative- and non-satisfaction cases are put up against the converse claim that content is just internal content.

The second claim of my argument is that, given certain common assumptions, this dialectical pattern exhausts what can be said about content. Abstracting entirely from the concrete issues, we have three notions: external content (what is 'out there'), internal content (what is 'in here') and content *simpliciter* (the significance of whatever it is we are discussing). There are also a few widely accepted assumptions: that most names, thoughts and experiences have some external content; that ambiguity and contextual dependence can be safely ignored; and that if content is internal then the external content must be determined by this internal content.

Given all this, I claim that what has been said in each domain is *all* that can be said. There are two views available: that content is external content, and that it is internal content. I'll show that the cases discussed in the literature cover all the kinds of cases which could be brought up as counterexamples to either view.

Against the view that content is external content, three kinds of counterexamples are possible: pairs of cases with clearly identical content but different external content; pairs of cases of with clearly different content but identical external content; and clear cases of content without external content at all.

If there were any examples of the first kind, they would be cases of ambiguity or contextual dependence. For instance, the name 'Francis Bacon' might be thought to have a certain content, but this cannot simply be its referent, because there are several different people all called 'Francis Bacon'. This has been understood variously as a kind of ambiguity or contextual dependence. Either way, I

will set aside this kind of counterexample as those engaged in this debate have historically thought it safe to ignore such cases.

Examples of the second kind are instances of Frege's puzzle: pairs of cases which involve the same thing 'out there', but clearly differ in some important respect 'in here'. Examples of the third kind are failure cases: cases of clear significance, but without a worldly object to explain this.

These are all the kinds of counterexamples that could possibly be given. And all of them *are* given in the literature (given the assumption that ambiguity can be ignored). So the kinds of cases discussed in the literature exhaust the responses that can be made against a view that content is just external content.

The same is true of the converse position. Against the view that content is internal content, and with the assumption that then internal content would determine external, there are three kinds of counterexamples possible: cases of external content with internal content too broad to determine that thing uniquely; cases of external content with internal content too narrow to determine that thing at all; and cases of internal content (sufficient to determine some external content) but without external content.[33]

The first are just alternative-satisfaction cases: where the internal content is not specific enough to pick out the referent *uniquely* and another external object satisfies the internal content as well. The second are non-satisfaction cases: where the internal content doesn't pick out the 'right' object at all. Just as with the originally described alternative- and non-satisfaction cases, such cases are not mutually exclusive. I think the third kind of case is best understood as a non-satisfaction case too (perhaps as a limiting or degenerate version): the external content cannot satisfy the internal because there *is no* external content to do so.

We see the same pattern as before. These are all the kinds of counterexamples that could possibly be given, and all of them *are* given in the literature. So the kinds of cases discussed in the literature exhaust the responses that can be made against a view that content is just internal content too.

Now finally for the third step of the argument, that the truth of Unity is the best explanation for why the literature is comprehensive in this way. I argued in the second step that the cases given in the literature exhaust what can be said about content. My argument had two parts: first to establish what cases could be

[33]Following this schema, it might seem that there could be cases with no internal content but some external content, but such cases are impossible. Examples of this kind would be things like a perceptual experience which presented a certain real object but which did not seem to the subject to present any object at all, or of a name which referred to a certain person but with which the speaker associated no beliefs at all (not even that the person, whomever they were, was so-named). Clearly there are no such cases.

given, and second to show that just these cases *are* given. The argument for the first part was very schematic—I deliberately did not appeal to any claims about language, thought or perception in particular. All that was necessary was the model of external and internal content, and a few widely shared assumptions about it.

So not only do the three literatures, on language, thought and perception, say similar things, but they similarly exhaust what can be said about content. This is a much stronger similarity than that relied upon in the epistemological argument. My claim is that the best explanation for *this* similarity is Unity. It would be a truly bizarre coincidence if they each fulfilled this structure for other reasons.

This argument has an important corollary for later chapters: I have argued that we should believe Unity because the same dialectic is exhausted in three different domains. This dialectic includes four puzzles to do with the relation between external and internal content and the objects of the world. So if Unity *is* true, then in all three domains, the nature of content is just whatever explains or solves those four puzzles (and perhaps others, if the simplifying assumptions are dropped). This idea is the basis of my argument for the view I present of the nature of content, in Chapter 4.

Two final caveats. First, it might seem that if my argument succeeds, and the existing literature really does exhaust what can be said about content, then there is little point in writing a thesis on the topic! But the arguments above relied heavily on widely shared assumptions, which I do not hold. The claim that ambiguity can be ignored is effectively dropped in Chapter 3, and in Chapter 5 I argue positively that internal content need not determine external, even when both occur. My project advances away from the existing dialectic, not within it.

Second, I said at the beginning that my goal was to provide a good reason to believe Unity. I believe I have done this. But this is a fairly humble project: there being a good reason to believe it is consistent with Unity actually being false, with my reason being defeasible (or defeated), with there being good reasons to deny it, and with there being good reasons to believe it besides the one I have given. I did not set out to provide a knock-down argument to take on all comers, and do not believe I have done so.

# Chapter 3

# The Bearers of Content

So far I have motivated the claim that the objective significance of language, thought and perceptual experience is the same. The previous chapter attempted to give positive support to this claim. But this claim does not entail any particular view of exactly *what entities* have this significance—which 'bits' of language, thought or perception? Certainly there are many entities across all these domains which everyone will agree do not—blur is a perceptual phenomenon, and rising inflection for question sentences is a linguistic one, but neither seem to have anything to do with how language or perception relate to particular objects. Whether my experience is blurry, or whether my utterance includes rising inflection, do not affect either *whether* I perceive or discuss some particular object, nor *which* object I perceive or discuss. To see this, consider any pair of experiences or sentences which differ only in being blurry or not, or having rising inflection or not, and observe that both members of the pair relate to the same object(s) (if any). So if blurriness and rising inflection do not matter, the question here is which entities or phenomena *do*—if one relatum is always the worldly objects, what are the others?

I will use the term 'refeme' for the bearers of content, whatever they are. It should be clear from the previous chapter that content has some important relation to reference, and the -eme suffix is used in linguistics for the basic units of some category or type of behaviour: refemes are the basic bearers of referentially-significant content. This is supposed to be a philosophically neutral term—someone who thinks that thoughts are the bearers of content would claim that all refemes are thoughts, not that thoughts *as opposed to refemes* are the bearers of content. Hopefully, using a new term will also discourage any prior bias on these questions creeping into how I now formulate them.

It bears repeating that my project is only concerned with the bearers of a certain kind of content: that which names, singular thoughts, and perceptual experiences have been thought to share, and that refemes are whatever bears this

kind of content. There are a huge variety of linguistic, cognitive and perceptual entities: sentences, clauses, definite and indefinite descriptions, stories, particle terms, syllables, concepts, mental files, memories, qualia, dreams, phosphenes, phantom itches, etc.. The term 'content' has been used in various ways, and often to include the kinds of significance that any or all of these might have. It is perfectly intelligible to speak of the content of a sentence or a dream. But not all of these are *refemes* as I use the term. If Unity is true, then there is a subset of the linguistic entities, a subset of the cognitive, and a subset of the perceptual, such that any element in the union R of these subsets can bear a certain *kind* of content—the kind of content that names or singular thoughts have been thought bear. This is different from the kind of content that elements outside R might bear: names do not have truth values or imperative force (though some sentences do) and singular thoughts do not have narrative structure (though stories and dreams do). R is the set of the refemes.

This chapter is concerned with the question of what, metaphysically, refemes are. §3.1 explains more fully the significance of this question in the overall project. The question only arises if we take Unity seriously, and wish to understand its consequences with care. Since mine is the first explicit discussion of whether Unity is true, this is also the first investigation of the nature of refemes in these terms. There are widespread assumptions about the nature of linguistic, cognitive and perceptual refemes, and there has been some work on each of these areas respectively (which I discuss below), but nobody has compared the nature of refemes *across* these domains in light of Unity. §3.2 covers some preliminary clarifications, and in §3.3 I discuss the metaphysics of linguistic, cognitive and perceptual refemes separately. That is, I argue for views of what kinds of mental, linguistic, and perceptual items bear content. I show that the linguistic items are utterances (§3.3.1), the mental items cognitive episodes—tokens of concepts or mental files (§3.3.2)—and the perceptual items are 'object files' (§3.3.3) before some concluding remarks looking ahead to the next chapters.

## 3.1 Significance of this question

The question of the metaphysics of refemes is significant for at least two reasons.

First, it is not immediately obvious what refemes *within* any of these domains might be, but it must become at least somewhat more clear before good examples of the relevant phenomena can be generated for further inquiry. Most philosophy, including nearly everything cited here, proceeds by considering and comparing clear cases, edge cases and unusual or different cases of the relevant phenomena,

and trying to build good theories to encompass all and only the right cases. It is important then that we know *what* the consistuents of cases are supposed to be.

Take language: it is clear that something-to-do-with-names has something to do with objects. But exactly what? The same name can relate to the same object even though it is used on different occasions—we might thus conclude that really it is the *name* which is significant, not any one of the utterances. But by a similar token, the same name can be used to refer to entirely different objects on different occasions of use, either because it is borne independently by them both (I know several Alexes) or because a single pattern of use has switched from one to the other (Evans' 1973 'Madagascar' case).[34] We might thus conclude that it is really the *utterances* that are significant, as many causal theorists of reference have in order to answer Evans' challenge. This will have implications for the analysis of the puzzles which motivate the theory: if *names* are refemes then Frege's puzzle is about the difference in content between *two* things (the name 'a' and the name 'b'). If *utterances* (or inscriptions or other kinds of token) are refemes, then it is about a contrast between the similarity of the content of two tokens (those in the first sentence) and the difference of the content of two further tokens (those in the second)—a total of *four* things. Or take thought: philosophers often talk of *thoughts* as having content of the kind I am interested in, but Sainsbury (2010) argues carefully that really *concepts* bear content and these are 'used in' thoughts. Recanati (2012), Goodman and Genone (2020) and other theorists of 'mental files' have thought similarly. Finally, most philosophers of mind write of *perceptual experiences* having objective content. But for Pylyshyn and collaborators and other theorists of 'object files' (or 'FINSTs' in Pylyshyn's case), these have the objective content, and there can be up to five of them for every experience. (see e.g. Pylyshyn and Storm 1988; Scholl and Pylyshyn 1999; Scholl, Pylyshyn, and Feldman 2001; Pylyshyn 2001) So for each domain (linguistic, perceptual, mental) there is first the question of what entities bear content within that domain. Answers to this first question—separate answers for each domain—are developed in §3.3.

Second, these domains are massively disparate and there are significant disanalogies between the kinds of things generally assumed to bear content *in each*. The received view is that perceptual experiences, thought episodes and names all bear content, but these are metaphysically very different kinds of thing. Experiences and thoughts are temporally extended events, but it is not obvious that names are (and if they are, as in Kaplan's (1990) stage/continuant model, then they are composed of many disjoint episodes of utterance, unlike

[34]This is a common view. See Sainsbury (2014) for a view of names which excludes this possibility.

experience or thoughts[35]). Searle (1983 e.g. Chapter 1, Section iv) recognises a similar disanalogy between acts and states.

There are also apparent differences in the kind of content borne. One can have a thought or perceptual experience of more than one object at once, but names (or their utterances) are generally thought to refer to at most one thing at a time.[36] Similarly *which object(s)* a single perceptual experience concerns might change over time, though the same does not seem true of thoughts. There is a sense of 'perceptual experience' in which a subject could have a single experience over time, initially as of one object, which then passes away and is replaced by a noticeably distinct object. Although it seems to the subject that the object at the beginning is different from that at the end, it nonetheless seems that they have had only one experience. The same is not true of thoughts: if it seems to a subject that they have had thoughts about two different objects over a span of time, it cannot seem to them they had only one thought, which was first about only one and then about only the other. Names are more like perceptual experience in this respect (in Evans' 'Madagascar' case one name relates to different objects at different times) utterances are more like thoughts.[37]

Finally, there are apparent differences in how different kinds of refemes might have their content fixed. Names and utterances are much more public than either thoughts or perceptual experiences. If names or utterances have content, then plausibly it is fixed publicly too. It is possible to use a name incorrectly because there are rules of use independent of each of one's utterances; it is not possible to have the wrong thought or perceptual experience. (This is effectively Searle's (ibid.) distinction between world-to-word (or world-to-mind) and word/mind-to-world direction of fit). This list is probably not exhaustive.

These disanalogies, and any resolution they have, will impact any unified theory of content we can give. For one, it isn't clear how the same kind of content can be borne by drastically metaphysically different entities; we might think a sculpture is beautiful, and a mathematical proof also beautiful, but it seems that the beauty of each is rather different because the two things are so different. Perceptual experience is very different from language, so even if there is some sense in which they are functionally similar, we might worry that they can't really both bear the same kind of content—just as beauty is not the

[35]One interesting exception to this disanalogy is a comparison between names like 'Vulcan' and recurring hallucinations (i.e. the recurring hallucination itself, not any one episode within the recurrence). See Johnston's (2004 pp. 162–3) comparison of 'Vulcan' and John Nash's hallucinated 'friend' Charles Herman in *A Beautiful Mind*.

[36]Or at least, the kinds of multiple-object relations names bear are not like those which experiences or thoughts bear. Names are often to said to be 'ambiguous' between many bearers, but a visual experience of my cup next to my laptop is in no sense *ambiguous* between them.

[37]Though see Sainsbury (2014).

same for the sculpture and the proof. Many philosophers who believe that the content of language, perception and thought *is* similar also believe that what content is borne by some refeme of one kind can be determined by its relation to some refeme(s) of some *other* kind. For example, Sainsbury (2005) thinks that some cases of linguistic reference have their referent determined by their relation to mental tokens of intention (p. 107) and Devitt (1981 p. 133) appeals to perception and object-directed thought to explain the initial grounding of linguistic reference. To spell out a view like this in detail will require a theory of exactly what the things with content *are*, and how they can relate to each other (causally, iconically, syntactically, etc.).

So there are two related issues. First, it is not clear what kinds of things refemes *are* in any relevant domain. Second, those which are generally assumed to be refemes appear to be drastically different from each other, which should give pause to anyone positing a unified theory of their content.

There are various strategies for resolving these issues. One might accept the received view (names, experiences and thoughts are refemes) but explain away the apparent dissimilarities. Or one might accept the received view, accept the dissimilarities but argue they do not get in the way of a unified theory of content, pairing the eventual theory with an explanation of how these very different kinds of thing can have the same content, though they are metaphysically and functionally very different.

A different strategy begins by rejecting the received view, or at least not taking it for granted. The disanalogies of the second issue are only concerning under the initial assumptions that these (names, perceptual experiences, thought episodes) are the kinds of thing which bear content. But the first issue was that it is not at all obvious whether these *are* the kinds of thing which bear content. I adopt this strategy.

I will argue for positive views of which linguistic, cognitive and perceptual entities bear content and it will turn out that they are not so dissimilar after all. The bearers of linguistic content are not names but utterances, of cognitive content not thoughts as traditionally conceived but tokens of concepts or mental files, and the bearers of perceptual content not experiences but *object files* (these terms will be defined in more detail below). Giving these clear definitions deals with the first issue. The second falls away by default, since these are metaphysically and functionally very *similar* kinds of thing.

## 3.2 Preliminary Clarifications

Two brief clarifications before my arguments. Some philosophers have made fairly substantial assumptions about the relations between answers to these questions for different domains, and often used these assumptions in discussing others' views or formulating their own. Following a careful discussion of what makes for singular and general terms, Crane (2011b p. 22) says 'It is widely accepted that just as there are general and singular terms, there are general and singular thoughts'. Crane is careful to defend many claims in the paper, but important here is his implicit claim that, if there is a kind of mental entity which admits of a singular/general distinction like terms do, then it is thoughts. This assumption goes undefended. Jeshion introduces her notion of singular thought in a similar way. When presented with a rose, she says, 'I can visually attend to the rose itself and think *that is lovely*, where "that," as it functions in my thought, refers deictically to the object I attend to. (2010a p. 1, italics in original) Jeshion assumes that there is a sense in which a linguistic term ('that', and later 'it') functions in a thought.[38] Almog (2005) assumes (or attributes to Frege the assumption) that if a unified description of thought and language is to be given, then the same kinds of content must be attributed to *minds* as to *sentences*.[39]

Since claims like these are exactly the topic of this chapter, I have tried to avoid such assumptions: my arguments about the nature of refemes in language, thought and perception proceed independently of one another.

A second common assumption is that the only sentences or thoughts worthy of discussion are ones with truth conditions (sometimes called 'assertive' or 'declarative' sentences/thoughts). Crane (2011b) goes on 'The content of a thought is normally taken to be a proposition—that is, something assessable as true or false—and that is how I shall understand it here'.[40] Similarly Russell (1905) begins "On Denoting" with a list of what he calls denoting phrases, and seems to offer a theory of these. But really he gives a theory of truth-apt sentences which *include* denoting phrases. The move from phrases to sentences is carefully explained (p. 480), but the restriction to only truth-apt sentences is not, though it plays a significant role in the argument and final view.

[38]Of course, one might say that Jeshion is only writing loosely, and really meant to discuss something like how terms like 'that' function in *expressing* some thought. Even if this was the intended meaning, surely such loose writing is to be avoided when the topic at hand *is how thought functions*!

[39]Worse than merely assuming, Almog seems confused on this point. He is clear early on that the important comparison is between sentences and *minds*, but on page 494 moves tacitly to a comparison between sentences and *thoughts*. He offers no justification for either comparison.

[40]In other work, Crane notices the danger of this assumption and carefully holds back from it. (e.g. 2009 p. 478)

The danger of this assumption is that at least two kinds of thing are common to all the cases considered; first that they contain content (in the sense I am concerned with), and second that they are truth-apt. Thus, an analysis of the former may accidentally rely on features of the latter. Of course, this is only a *possibility*—it may well be that Crane's or Russell's analyses *can* be applied to non-truth-apt phenomena just as well.[41] But the possibility is entirely removed in the first place if the theory is constructed with a wider variety of examples in mind. I will try to adopt this latter strategy.

The considerations which led to my present project are clearly not limited to truth-apt phenomena. I can *wonder whether* Hesperus is Phosphorous. It seems clear to me that one person thinking (truth-assessably) that Hesperus is bright then wondering (non-truth-assessably) whether Hesperus is Phosphorous must exercise on two occasions *one* capacity to think about Hesperus (compare Evans 1982 p. 75, discussion of the 'generality constraint', though Evans too deals mostly with truth-apt content). Similarly someone who asks 'is Hesperus visible?' must be using 'Hesperus' in the same way as someone answers 'Hesperus *is* visible', otherwise the latter would not be an answer to the former. And one can just as well instruct someone 'Shave the King of France's head!' as assert whether he is bald, though Russell's analysis can only handle the latter case. I have thus avoided this assumption too, and tried to consider a broader array of examples.

## 3.3 Refemes…

I argue for theories of the nature of refemes for each domain separately. That is, I argue for views of what kinds of mental, linguistic, and perceptual items bear

[41]Grice (1991) raises a similar concern, which he attributes to Searle (Grice gives no citation, but Searle's 1969 Section 7.1 is a likely source). Grice is initially unimpressed by the application of Russell's paraphrasing method to commands or (especially) questions; he thinks the paraphrases are 'about' things which the originals were not. Chisholm (1977 pp. 171–2) also suggests that there are some clearly truth-apt sentences for which Russell's analysis fails. These criticisms aside, how much all this impacts Russell's view depends on what one takes to be essential to it. Clearly a paraphrase analysis in the spirit of Russell's ideas could be given of non-declarative sentences, such that 'is the King of France bald?' would correspond with something like 'if there is exactly one thing which is the King of France, is it bald?'. But Russell himself is very clear (e.g. on page 481) that the analysis for the sentences he considers is to be carried out in *truth functional terms*—he says that 'C(a man)' represents *any* sentence containing the denoting phrase 'a man' and that 'C(a man)' is equivalent to "C(x) and x is human' is not always false'. If the view proper is only the general idea of paraphrase analysis, then Russell is only guilty of providing rather limited examples. If the logical machinery is also part of the package, then the view is simply false, and Russell failed to detect this because of his limited examples. But precisely *because* Russell only considers declarative sentences, it is impossible to know which is the intended interpretation. Either way, is it not clear that one should not rely entirely on declarative sentences?

content. The arguments are separate because I don't want to assume what I am setting out to show—that refemes in each domain are unproblematically similar (or equivalently that they are *not* problematically dissimilar).

I show that the linguistic items are utterances and similar tokens like signings (§3.3.1), the mental items cognitive episodes—tokens of concepts or mental files (§3.3.2)—and the perceptual items are 'object files' (§3.3.3).

The arguments are separate, but the method is similar in each case. Recall that the content borne by refemes is just that which, within philosophy, allows us to explain or understand the four puzzles of the first chapter in a unified way. To do this, we must be able to properly describe those puzzles in the first place—the puzzles must be *sensible* before they are *soluble*. In each case, I argue that if we are careful, then the puzzles are best understood if we take refemes to be of the metaphysical kind listed above (utterances/tokens/object files).

Call this the *methodological argument*. Alongside this main line of argument, I have also included a couple of other, more specific considerations in favour of my views at various points.

### 3.3.1 ...Linguistic

There are two views one might reasonably hold concerning the metaphysics of linguistic refemes. Language includes names, descriptions, demonstrative terms and other kinds of *words* or *phrases*, distinct from any event of *use*.[42] The 'official' view, (hereafter the *terms view*) outwardly expressed or assumed by many, has it that these terms or phrases bear content. McGinn (1981) begins 'Reference is what relates *words* to the world of objects on whose condition the truth of sentences hinges'. (p. 157, my emphasis) Sainsbury (2005) develops his approach to a theory of meaning entirely in terms of names (see e.g. his axioms of reference on p. 73). Frege (1948) takes sense and reference to be properties of 'signs' and Kripke's (1980) discussion is couched in terms of names.

Each of these terms or phrases can be *used* in various ways—uttered, written, brailled, signed[43] etc. The other view is that these uses are the bearers of con-

[42]My presentation sidesteps the difficult question of what exactly these words or phrases are, and the related one of the conditions under which two tokens are of the same word or phrase. It may even be that 'word' is the wrong category, and a discussion of morphemes word stems would be more appropriate (the stem is the common part of, for example, 'Heagren', 'Heagren's' and 'Heagrenian'). The central idea is just that there is something common to several utterances, and distinct from any one of them, and that this is the thing which bears content. My discussion of the terms view is unaffected by issues of word or stem metaphysics or individuation, and 'term' can serve as a shorthand for whatever the general content bearer is, so like all others cited here I will leave the issue aside.

[43]Spelled out in a signed alphabet. Signs in a full sign language like British Sign Language are 'utterances' of words in a different language.

tent. Terms like 'use', 'utterance' and 'signing' display what Grice (1957) called 'act/object ambiguity'. On any occasion of use there is an act of producing an utterance (or sign, etc.), and there is the utterance (sign, etc.) *which is produced*; these are at least conceptually distinct. In spoken and signed language, the distinction is easy to miss, because the act of production coincides with the product. Some might even say that the distinction is merely conceptual, and the act of producing just is the thing produced.[44] The distinction has thus been safely ignored in much previous work (including Grice's own). Nonetheless, it is important, and it is more obvious in cases of writing or other recorded language. If I write a note including someone's name, the act of production is over quickly, but the inscription of the name might stick around for years and be interpreted only later. Surely it is the inscription (the 'product') which has the content, and not the act of producing it. If not, then it would mysterious what was being interpreted when the message was eventually read. Presuming that written and spoken language are roughly similar in this regard, if there is a metaphysical distinction between utterance or signing acts and their products, then it could only be the products which bear content. The second view then is that these products are the bearers of linguistic content. I call this the *uses* view, though bearing in mind that 'use' is understood on the product interpretation.

The uses view is supported implicitly by much recent theorising about reference: many sophisticated modern theories require or assume (or occasionally, positively argue) that utterances or other such tokens bear reference. (Sainsbury 2005 Chapter 3; Michaelson 2023a)

I will argue against the terms view, and thus in favour of the uses view.

A first and simple argument is that the terms view is only plausible under the simplifying assumption (noted in the previous chapter) that each name has at most one referent, but this assumption is false. Take a failure case involving the name 'Vulcan'. The terms view claims that what makes the case a failure case is a fact about the content of the name 'Vulcan'. This is plausible if we only consider the content of 'Vulcan' with respect to Le Verrier and his apparent scientific discovery. But in fact 'Vulcan' is ambiguous—a town in Alberta, an Irish political magazine, and several novels and albums are all called 'Vulcan', and all exist. If it is just the name 'Vulcan' which bears content, then the case is at best a partial or ambiguous failure. Clearly it isn't, but the terms view can't support this conclusion (though see Sainsbury 2014 for a more sophisticated proposal). Similar concerns apply to the other kinds of case: if 'Hesperus' is ambiguous then 'Hesperus is Hesperus' might not be trivial (it might not even be true!) and it might turn out that the alternative satisfier in an alternative

[44]Hanks (2015) defends a view along these lines but for the bearers of truth and falsehoods.

satisfaction case really is one of the referents of the relevant name. Conversely it is plausible that uses of names have only one referent, though different uses of the same name might refer to different things, so the uses view can recover these cases without simplification in a way the terms view cannot.

Now the methodological argument, as applied to linguistic refemes. It is uncontroversial that Kripke (1980) advances arguments against descriptivism. It is less clear exactly what the descriptivist view is which these arguments are designed to attack. Nonetheless, arguments of this type have been taken to be important moves in the debate over the nature of (linguistic) content. I will argue that, if we are to take them to have any dialectical force, cases of the type Kripke presents must be interpreted as concerning *uses*, not the names or terms independent of any such use.[45] Since refemes are just whatever-bears-content, and content is whatever-makes-sense-of-puzzles-like-these, this is sufficient to show that linguistic refemes are uses. This is the methodological argument described above.

I consider alternative-satisfaction cases first (using Kripke's 'Feynman' example), then non-satisfaction cases, and finally show that the same methodogolical argument applies to failure cases, with similar results. In each case, I argue by assuming my conclusion is false (i.e. that the cases are best understood as dealing with terms, not utterances) and showing that this interpretation fails.

Take first Kripke's alternative satisfaction case of 'Feynman'. Kripke says:

> Consider Richard Feynman, to whom many of us are able to refer. He is a leading contemporary theoretical physicist. Everyone here (I'm sure!) can state the contents of one of Feynman's theories so as to differentiate him from Gell-Mann. However, the man in the street, not possessing these abilities, may still use the name 'Feynman'. When asked he will say: well he's a physicist or something. He may not think that this picks out anyone uniquely. I still think he uses the name 'Feynman' as a name for Feynman.
>
> (Kripke 1980 p. 81)

Suppose the terms view is correct: then there is a property (being a famous physicist) somehow associated with the *name* 'Richard Feynman', which is satisfied by something other than Feynman himself (Gell-Mann above, and many others besides) and importantly *there are no other properties similarly associated with the name which distinguish Feynman from Gell-Mann*.

[45] In Kripke's puzzles these uses will be *utterances*, though manual signings could stand in just as well.

This is in fact false. The name is a public entity, used by many speakers (at many times, in many different communities etc.) and any property associated with it is associated with it *relative to some speaker* (i.e. *by* some speaker)—no properties are associated with the name *simpliciter*. In other words, though the name itself is public (two people might utter the same name on different occasions), any association between name and property must be private to an individual. To see this, consider how the relevant property (which is satisfied by Feynman and other individuals besides) might vary between speakers.

Suppose Feynman gets mixed up in a robbery investigation and ends up as one of the suspects. The officer in charge has no interest in physics and has never heard of him before, but still uses the same name[46] '[Richard] Feynman' to discuss him, having seen it on a list of suspects (and suppose that at least one of the other suspects is similarly unknown to the officer—they are not all regular offenders). Like the man in the street, the officer cannot pick out Feynman by any uniquely-held property, because they know just as much about the other suspects as they does about Feynman. When asked about Feynman, the officer will say: well he's a suspect in the case or something. Just as with the man in the street, this property is not uniquely held by anyone, and (again like the man in the street) the officer may even know this themselves. And yet, the officer still manages to talk about Feynman when they say things like 'let's interview Feynman first'. The officer knows nothing about notable physicists, and the man in the street nothing about the robbery case, but they still use the same name to discuss the same person.[47]

To make sense of Kripke's case, the refeme must be associated with some property which is not uniquely satisfied. The case above shows that such associations are specific to speakers. So, if we wish to make sense of the idea that there is something linguistic which might be the bearer of content, and is associated with a property which is not uniquely satisfied, then the 'something linguistic'—whatever is—must be something specific or relative to *speakers*.

A reasonable candidate is *uses* (in speech, *utterances*). Uses by the man in the street are associated with the property *being a famous physicist* and utterances by the officer with *being a suspect*.

'Not so fast!' an objector might say—the argument so far does establish that linguistic refemes must be something specific to speakers, but utterances are very

[46]This is true even of relatively stringent theories of sameness for names, such as Sainsbury (2014). On Sainsbury's account we would say the officer and the man in the street use the same *specific* name (which on Sainsbury's account they do, since their uses are both part of the same causally-constituted practice).

[47]Searle (1969 p. 168-9) makes a similar point. He dismisses Frege's (1956) discussion since 'according to Frege, unless our descriptive backing for the name is the same, we are not even speaking the same language'.

specific—why not just accept the terms view and index names to speakers? We would say then that the-name-in-the-officer's-mouth (or the-name-for-the-officer, etc.) is associated with *being a suspect* and the-name-in-the-man-in-the-street's-mouth is associated with *being a famous physicist*.[48]

Such talk can be interpreted in two ways. Either 'in-someone's-mouth' is an adjective modifying the name under discussion (so we are not discussing a name, but a name-in-the-officer's-mouth) or it modifies the association relation (so that the name is associated-with-respect-to-the-officer with the property).[49] It is not clear to me what a name-in-a-person's-mouth is supposed to be, and I am not aware of any well-developed account of this, so the first interpretation seems unlikely. The second interpretation seems simply to collapse into the uses view.

Moreover, the same type of argument as above—varying the property associated with a name as the speaker varies—can be run to show that the associated property can vary over time for the same speaker, so merely indexing the property-association to a speaker is not enough. This shows that we cannot simply index names to speakers (as the two interpretations above do), we must index to speakers *and* points or spans of time. Consider a extension of the officer/Feynman case. The police officer has a chequered history: they were once a keen student of physics, but the pressures of grad school were too much and drove them to addiction, which eventually caused them to forget everything they knew about it. During their very earliest experiences with physics (say, as a teenager in school, call this time T1) if asked about Feynman, they would have said: well he's a physicist or something. Later on, they will know more about his work and perhaps be able to describe him uniquely in terms of his theories (T2). Finally, following their addiction and consequent amnesia they become a police officer and their only semantic interaction with Feynman is as a suspect in a case (as above—this is T3). The person at T1 knows nothing about the case at T3 (in part because it hasn't happened yet!), and the person at T3 nothing about notable physicists, but they still use the same name to discuss the same person.

Once again then, if we wish to make sense of the idea that there is something linguistic which might be the bearer of content, and is associated with a property which is not uniquely satisfied (i.e. if we wish to take non-satisfaction cases seriously), then for a given speaker the 'something linguistic'—whatever is—must be something specific or relative to *times*. Combining this with the earlier

[48]Field (1973), Braun (2005), Gómez-Torrente (2019), deRosset (2020), and Koch and Wiegmann (2022) all write in such terms. On page 82 Kripke (1980) discusses the semantic properties of a name 'for a speaker'.

[49]One way of cashing this out would be to claim that the association relation really has three relata: the name, the property, and the speaker (in this case the officer); another way would be to claim that there is a separate association relation for each speaker (each such relation being binary). This difference should not affect my discussion.

conclusion we get that the something—whatever it is—must be specific or relative to *speakers and times* (or speakers-at-times).

Again, a reasonable candidate is *uses*—uses by the man in the street and the person at T1 are associated with the property *being a famous physicist* and uses by the officer at T3 with *being a suspect*. This view also accounts for the very rapid changes in associated property for the person at and around T2. As a student of physics (perhaps even studying Feynman's work) they will learn lots of things about Feynman over a short span, and so even two adjacent utterances may be associated with quite different properties; consider for example the utterances concerning Feynman made immediately before and immediately after a lecture on his early work, which will significantly expand the student's beliefs.

The argument so far has been entirely in terms of alternative-satisfaction cases. The same method can be applied to other cases too.

Consider a non-satisfaction case. Recall my case of Alice and the history of mathematics. Alice associates, mistakenly, the name 'Andrew Wiles' with the property *prover of the collatz conjecture*. When asked about Wiles, Alice will say: well, he's the one who proved the Collatz Conjecture.

Suppose the terms view is correct: then there is a property (*being the prover of the collatz conjecture*) somehow associated with the *name* 'Andrew Wiles'.

But again, this association holds only relative to certain speakers. Bob has also heard of Andrew Wiles, and also holds mistaken beliefs about him, but unlike Alice associates the property *being the prover of the Goldbach conjecture* with the name (the Goldbach conjecture also remains unproven). When asked about Wiles, Bob will say: well, he's the one who proved the Goldbach Conjecture. But presumably both Alice and Bob manage to talk about Wiles, though they believe and say false things about him.

This is clearly a non-satisfaction case: the only properties associated with the name (within the tiny universe we are considering) are ones not in fact satisfied by Wiles. But in virtue of which property are we to say it is a non-satisfaction case? Alice doesn't believe Wiles proved the Goldbach Conjecture, and Bob doesn't believe he proved the Collatz Conjecture. Once again, the features of the case which make it the kind of case it is (a non-satisfaction case) can vary between speakers, so whatever the thing is which is associated with the property must also be speaker-specific. And as above, a similar case could be constructed with one speaker holding various different beliefs about Wiles at different times; initially only that he proved the Collatz Conjecture, then only that he proved the Goldbach Conjecture. This too constitutes a non-satisfaction case, but again there is no one property which makes this so. Just as above, this demonstrates that the association between refeme and property must be indexed to a time as

well as a speaker, and as before, taking refemes to be utterances allows us to make sense of this (and taking refemes to be terms would not).

Finally, take the case of empty names. These cases work because the empty names (or their utterances) appear to function similarly to referring names but in fact do not refer: (utterances of) sentences like 'Vulcan is a planet' or 'Vulcan is a telescope' seem meaningful, and the former are in some sense more appropriate than the latter. With a referring name, this behaviour is explained by the association of (utterances of) that name with properties which the utterance's/name's referent has; it is more appropriate to say that Hemingway was a writer than that he was a baboon, because he, Hemingway, *was* a writer, and *was not* a baboon. So (utterances of) referring names function this way because they are associated, by reference, with certain properties (somewhat like the alternative satisfaction case). This model also applies to empty names: (utterances of) such names are associated with properties, and it is more appropriate to say things which affirm this relation than deny it; 'Vulcan is a planet' is in some sense better than 'Vulcan is a telescope' because the name 'Vulcan' has historically been more generally associated with being a planet (or perhaps with referring to something which is a planet) than with being a telescope.

Exactly as in the alternative satisfaction case, *which* properties are so related to names/utterances can vary between speakers and between times (and exactly as before, this shows that the association cannot really be with *names*, but instead with *uses* of names). Take inter-speaker varation first: two hungover students attend an early morning lecture on philosophy of language, but miss most of the details. One leaves believing that Vulcan is a planet (but has forgotten Le Verrier) and the other that Vulcan is something Le Verrier discovered (though they forget what). Now suppose that the terms view is correct—are we to say that the (empty) name 'Vulcan' appears meaningful because it seems appropriately used in sentences such as 'Vulcan is a planet' or because it seems appropriately used in sentences such as 'Vulcan was discovered by Le Verrier'? Clearly neither is straightforwardly correct. Sentences of the first kind only appear meaningful for the first student (and for anyone else who has the appropriate Vulcan/Le Verrier beliefs), whereas sentences of the second only appear appropriate for the second (and for anyone else who has the appropriate Vulcan/planet beliefs). Once again, to make sense of the puzzle, we must index content to a speaker.

Now consider variation for the same speaker over time. A child learns the name 'Santa Claus' and believes that it refers to a jolly old man who brings gifts at Christmas. As they grow older, they lose this belief, but come to believe instead that 'Santa Claus' refers to an ancient saint (and let us suppose that in fact neither of these is the case). Perhaps in their dotage they even forget that

they ever believed in the jolly old man. Should we say that for this speaker, 'Santa Claus' is associated with the property of being a jolly old man who brings gifts, or with being a saint? Clearly neither is straightforwardly correct—the first is right for the speaker *at a certain time* (or span of times) and the second for the same speaker *at a different time* (or span of times).

Another way of putting all this is that one of the above formulations was wrong: it is *not* the case that the *name* 'Vulcan' is associated, *simpliciter*, with the property of being a planet (or being a planet between Mercury and the sun, etc.), because some people, at some times, *don't* associate the name with that property. As with the alternative-satisfaction and non-satisfaction cases, if the empty names case is to have the dialectical force it is supposed to, it must be treated as dealing with *uses*, not *names*.

So, to properly state three of the cases which originally motivated my notion of content for language (alternative satisfaction, non-satisfaction, and failure/empty names), it is much more straightforward to take the bearers of content to be uses, not the names used (and as noted above, in this regard there is no significant difference between utterances and other token uses such as inscriptions or signings).

Of course, this is not an entirely decisive argument in favour of the uses view. A defender of the terms view might continue as above: they might say it really is *terms* which are the bearers of content, but the terms (or their relation to the content they bear) must be indexed to both speakers and times. What has content then is a name or other term *for* a speaker, *at* a time. This sort of view is suggested by occasional discussion of the meaning of a name 'in someone's mouth'. Again, I don't know exactly how this view would be made properly specific, and without more detail it is difficult to see how far this would really depart from the uses view anyway.

### 3.3.2 …Cognitive

There are four views one might reasonably hold concerning the metaphysics of cognitive refemes.

1. That they are complete cognitive episodes with complete contents (these are often propositions of some kind). (Crane 2011b p. 22; 2013)
2. That they are the contents *of* such episodes, as might be instantiated or thought on more than one occasion. For example, one who believes this, and believes that the contents of thought-episodes are propositions, will believe that cognitive refemes are propositions (Jeshion (2010a) presents Russell (1970; 1905) as having held this).

3. That they are concepts, mental files, or other such entities, relating to individuals, and deployable/tokenable on more than occasion. (K. A. Taylor 2010; Sainsbury 2010; Recanati 2012; see Goodman and Genone 2020 for more on the relation between mental files and singular thought)
4. That they are tokens or 'activations' of such concepts or files. (K. A. Taylor 2010; suggested by comments in Recanati 2012 pp. 96, 120)

(Taylor's view is nuanced, views 3 and 4 both play a role in his overall picture—see pp. 85–87 for details.)

These candidates can be grouped in at least two different ways. First, we could divide them into those which are repeatably instantiable (2,3), and those which are not (1,4). Contents of thoughts, concepts and mental files can all be tokened/activivated/used more than once over time. The tokens/activations/uses themselves cannot. This distinction is the same as that considered in §3.3.1 on linguistic refemes. In the first part of this section (§3.3.2.1) I'll argue that cognitive refemes must be of the non-repeatable type (just as I did with linguistic refemes). I employ the same kind of methodogical argument as in the linguistic case. This establishes that cognitive refemes must fall into categories 1 or 4.

Second, we could divide them into those which can relate to more than object at once (1,2) and those which cannot (3,4). Consider an episode of thinking *that man stole my wallet* (this example comes from Crane (2011b)). This episode, or what is thought in it, is about that man *and* about my wallet; more than one thing. Presumably thinking this thought requires at least two distinct concepts (or mental files, or something similar): one for that man, and one for my wallet. Each of these concepts is about only one thing. In the second part of this section (§3.3.2.2) I argue that cognitive refemes must be of the kind which are only about one thing; that they must fall into categories 3 or 4.

Taken together, these arguments show that cognitive refemes must fall into category 4: that they are tokens of concepts or mental files. As with linguistic refemes in the previous section, terms like 'token' and 'activation' are ambiguous: they could mean the acts or events of 'tokening', 'activating' or 'calling to mind' certain concepts or mental files, or they could mean the products of such acts. I am sceptical that there is a metaphysically relevant distinction here, but if there is then the bearers of content will be the products, not the acts. If an agent holds a thought about something in mind for some time, though the act of producing the mental tokens is only momentary, we would correctly say that they are still thinking about the thing after a short while; the content has persisted (along with the 'product'), though the act is over. Since there is no simple way to distinguish act from product (as with writing in the linguistic case) and the difference bears little on my arguments, I will mostly ignore this distinction in what follows.

#### 3.3.2.1 Repeatability

In this section I argue that cognitive refemes must be non-repeatable tokens, rather then the more general things which *are tokened* (concepts, mental files, etc.). As before, I'll argue by contradiction: suppose that cognitive refemes are something like concepts, mental files, or the complete contents of thoughts (e.g. propositions, or structured sets of concepts or mental files). What is important here is that any of these things can be tokened in thought more than once: the same proposition might be the content of more than one episode of thinking, as might a concept or structured set of concepts (assuming of course, a view on which such things are *ever* the contents of episodes of thinking).

One clarification before the argument: on some views of these tokenable entities, they are things which can be possessed and tokened by more than one agent. For example, some views have it that concepts are Fregean senses, which are public entities accessible by more than one agent. On these views, it is possible for different thinkers (at different times) to have thought-episodes with numerically identical content. If one of these views is true, then it would be possible to construct an argument to show that which properties are associated with some such entity (e.g. with a certain concept/sense) vary between thinkers; this would be exactly analogous to the argument I gave above showing that which properties are associated with a name can vary between speakers. This would show that the associations between cognitive refemes and properties (the associations necessary for making sense of alternative satisfaction cases) are thinker-relative; again, this is just like the linguistic case above. On any *other* view, the tokenable entities are not public, and therefore any associations must be thinker-relative anyway. Since views of the latter kind seem more popular and my argument with respect to the former kind would be *exactly* as in the names case, I will not develop this line of argument any more explicitly. I take it that this is enough to establish that cognitive refemes must be thinker-specific.

Now back to the main argument, concerned with variation in the properties associated with a tokenable entity by a single thinker over time. Consider the puzzles we began with. Once again, take an alternative satisfaction case first: I have a singular thought about some object O, that thought is associated[50] with a cognitive refeme, there is a property (or properties) associated[51] with that refeme which O in fact satisfies, *but* that property (or properties) does not uniquely pick out O because it is satisfied by something other than O as well.

[50]'Associated' is meant in a very general sense here. For those who believe (singular) thoughts *are* refemes, the 'association' will be identity, for those who believe (some of) the constituents of singular of thoughts are refemes the 'association' will be (partial-)constitution, etc.

[51]Here 'associated' is used in the usual way again.

We have supposed that the refeme in this case is a concept, mental file, proposition or tokenable entity, independent of any particular token *of* that entity. So a tokenable entity like this is associated with a property (viz. a property satisfied by O and at least one other thing).

To make this concrete, let us take Zarepour's (2018) case: as Jones walks by my house (at time T2), I have a singular thought about him, that thought is associated with my *concept of* him (or my mental file, or a structured proposition including him, or whatever), which is in turn associated with a property or properties. For simplicity, suppose that the properties are that he has brown hair, is 6' tall, and is wearing a leather jacket. It so happens that yesterday (T1) I had sat in exactly the same place, at the same time of day, in indistinguishable lighting, and seen John—Jones' indistinguishable twin—walk past my house. Exactly the same cognitive processes occurred, and as a result I formed a concept (mental file, structured proposition) of John, which was associated with having brown hair, being 6' tall, and wearing a leather jacket. For whatever reason though, I had partial amnesia overnight and have forgotten the whole thing.

This is clearly an alternative satisfaction case: the properties associated with the refeme are not sufficient to individuate Jones (or John for that matter) since they are satisfied by Jones *and* John.

If we wish to make sense of the case supposing that refemes are concepts (or mental files, propositions, etc.) then it must be the *concept* which is associated with these properties *simpliciter*.

This is in fact false. The concept persists between episodes of thinking in which it is used or deployed. Unless I have amnesia again I will retain my concept of Jones until the next day. Suppose that for some reason, Jones has access to video footage of my street and my window. He sees the footage of him walking past my house, and wants to know why I was watching him, so he comes to my door and asks. I recognise him and now (at T3) associate with the concept (the refeme) of Jones the property of having come to my door and asked why I was watching him. I never associate this property with the concept (refeme) related to John. So now the case is no longer an alternative satisfaction case: there is only one person who satisfies the properties I associate with the refeme (namely Jones).

Clearly, in order to make sense of this story as an alternative satisfaction case, we must discuss the properties associated with a concept (mental file, proposition, etc.) *at a point or span of time*, not those associated with it *simpliciter*; the latter version is problematic because which properties are associated with a concept can change over time (there is a clear parallel with my argument above

that names cannot be refemes, because the properties they are associated with change over time).[52]

This does not quite establish the claim I began with though: that *tokens* of concepts, mental files or structured propositions are the cognitive bearers of content. What has been so far established is that cognitive refemes must be time-specific. This is consistent with the claim that cognitive refemes are just concepts-at-times (or mental-files-at-times, propositions-at-times, etc.). This is clearly a much broader view: one might think that a cognitive refeme is just a concept over some span of time during which the cognitive agent does not alter which properties they associate with that concept. The next step of the argument is to demonstrate that such a view cannot be true, and it really must be *tokens* that are refemes.

Once again, the argument will be that is impossible to interpret alternative satisfaction cases with the dialectical force we need, unless we take cognitive refemes to be tokens. And once again, the argument will proceed by contradiction. So suppose cognitive refemes are just concepts (mental files, propositions) indexed to points or spans of time. And suppose I have two concepts, A and B. At time T, concept A relates to some individual—say Jones from the case above—and is associated with a property which is satisfied by that individual and others—say *having brown hair*. At the same time, concept B is associated with some individual—say Einstein—and is associated with a property which is satisfied by them uniquely—say *being the discoverer of special relativity*.

Let us stipulate that I have a singular thought at T about Jones, which makes for an alternative satisfaction case. What makes it an alternative satisfaction case? On the view under consideration it is not at all clear. At time T I have two concepts each associated with different properties, such that one concept (A) is satisfied by various different people, and the other concept (B) is not. From this description alone, it is impossible to say that an alternative satisfaction case occurs, because that is only true if we only consider concept A—concept B's associated properties *do* pick out a unique individual. To describe what happens as an alternative satisfaction case, we must discuss only concept A (and not B), but why privilege A over B? The answer of course is that A is the concept *which is tokened in my episode of thinking* (since the thought is about Jones). The thought counts as an alternative satisfaction case because it tokens the concept A (and not B) and A is associated with properties which do not uniquely pick out one person.

So: to interpret cognitive alternative satisfaction cases as having the kind of dialectical force that they must have (indeed, to properly describe them at all),

[52]Prosser (2019 p. 664) makes a similar point, though his discussion is rather compressed.

we must take it that the content in such cases is borne by tokens *of* concepts, mental files, structured propositions of other such entities, not these type-level entities themselves.

Of course, a staunch defender of a type-level view might hold that cognitive refemes are concepts (mental files, propositions) indexed to their own *tokens*. This is a similar move to one considered briefly above for names, where one could claim linguistic refemes are names or terms indexed to speakers and times. Here as there though, it is difficult to see how this view substantively differs from mine.

#### 3.3.2.2 Plurality

The previous section demonstrated that cognitive refemes must be tokens of something like concepts, mental files or propositions, rather than merely the concepts, mental files or propositions *themselves*. For a given cognitive episode, *the episode*, or some part or constituent of it, is the refeme, not the type-level concepts, mental files or propositions which it instantiates or tokens. In this section I argue for the latter option—cognitive refemes are tokens of concepts or mental files (or some other such thing) which are *used in* or *partly constitute* whole episodes of thinking; they are *not* the whole episodes themselves, or tokens of propositions.

Consider again Crane's example from above. I have a thought that *that man stole my wallet*. (Crane 2011b p. 23)

The propositional content relates to two objects: that man, and my wallet.[53] The concept or mental file *my wallet* relates to only one. I will argue that tokens of the latter are cognitive refemes, tokens of the former are not.

Once again, suppose otherwise, and consider whether the resulting view could make sense of the cases which motivated the notion of content to begin with. Then there would be cognitive tokens of things like propositions (or structured sets of concepts or mental files) which were instances of these cases. (from here on, I will just say 'propositions' as identical reasoning follows for these and the other cases). We will see that this is impossible.

As before, consider alternative-satisfaction cases. In such cases there is a refeme which in fact relates uniquely to one object, and is associated with some property or description, which is satisfied by something *other* than that object. But propositions are not, in general, uniquely related to one object. Consider

[53] 'Relates' here could be read as a binary relation, or partial-constitution, inclusion, or something else, depending on one's view of propositions. See King (2024) and McGrath and Frank (2024) for some options.

a cognitive token of the proposition that *the Moon orbits the Earth*.[54] Which object does this token relate to, with respect to which an alternative satisfaction case could be constructed? Clearly the question is misplaced—it relates to *two* objects: the moon and the earth. The structure of alternative satisfaction cases requires that there be *one thing* to which the refeme is uniquely related, and propositions don't work this way. Another way of putting this is to note that the relation between the refeme and the object in alternative satisfaction cases is often understood as *singular reference*. But asking what a proposition singularly refers to is a category error.[55]

The same line of reasoning follows for non-satisfaction cases. Take again a cognitive token of the proposition that *the moon orbits the earth*. In non-satisfaction cases there is a refeme which in fact relates uniquely to one object, and is associated with some property or description, which is *not* satisfied by that object. Which object is this proposition uniquely related to? Again, this question is simply misplaced.

Similar comments apply to failure cases. Suppose I think that *Hermes met Zeus*. Failure cases work because something behaves *as if* it relates uniquely to some particular object, but in fact does not (again, usually this relation is understood as reference). In this case, what is supposed to behave as if it relates to exactly one object, though in fact it does not? Certainly not the token of the proposition that *Hermes met Zeus*. This does not actually relate to any objects, since neither Hermes nor Zeus exists, but it behaves (in some ways) as if it relates to at least two: Hermes and Zeus.

So the following is true for all three of these cases: in order for the case (indeed, any case of that type) to play the dialectical role it is supposed to, the cognitive refemes cannot be tokens of *propositions*. To return to the earlier categories, by the same line of reasoning they cannot be tokens of structured sets of propositions, or the whole contents of episodes of thinking.

Instead, cognitive refemes are tokens—things specific to episodes of thinking—of *concepts*, mental files, or other such entities.

[54] I have moved away from Crane's example as it includes demonstratives and indexicality. Apart from the complications these introduce, his example should be subject to exactly the same arguments.

[55] There is a sense of 'reference' or 'aboutness' in which propositions *do* refer-to (or are about) things: the much looser sense that Crane (2009; 2011b; 2013) for example appeals to. But this is not the sense of reference/aboutness required of refemes in order to make intelligible the sorts of puzzles I am concerned with.

### 3.3.3 …Perceptual

Among philosophers of perception it is near-universally assumed that if there is any sense to talking of 'perceptual content', then the things which *have* this content are perceptual experiences—paradigmatically *visual* experiences. Such experiences are temporally extended episodes of sensing or perceiving. I will argue that, if we are to talk of the kind of content I am interested in, then this is wrong: the things which have content are *constituents* of perceptual experiences, but are *not* the (whole) experiences themselves. These constituents are what have previously been called 'object files' (Green and Quilty-Dunn 2021) 'perceptual pointers' or 'FINSTs' (Pylyshyn and Storm 1988; Scholl and Pylyshyn 1999; Pylyshyn 2001).[56]

To put the last paragraph in the terminology I have been using: it is widely assumed that perceptual refemes are perceptual experiences; I will argue that this is wrong, and that instead perceptual refemes are something like object files. Unlike the linguistic and cognitive cases, there is no ambiguity in the term 'object file'. It is the *files*, not the acts of producing them, which are refemes.

My argument has two parts: first, I show that perceptual refemes must play a certain sort of role, and that whole visual experiences do *not* play it. (§3.3.3.1) The methodology is similar to that in the sections on cognitive (§3.3.2.2) and linguistic (§3.3.1) refemes. Second, I show that object files, or something similar *do* play this role. (§3.3.3.2)

The empirical and philosophical resources I rely on below are concerned almost exclusively with vision. In principle, my conclusions could be generalised to other perceptual modalities as well. This is a common strategy and is necessary because I rely on ideas developed mostly in empirical scientific work, and that research programme has so far been limited to vision.

This strategy has occasionally been criticised for assuming, without good grounds, that such a generalisation is possible. Vision clearly presents us with objects, which seem to have properties, but it is not clear that the other senses do. Perhaps there are auditory objects, plausibly there are haptic ones, but it

[56]Pylyshyn (e.g. 2007) writes of 'FINSTs', short for 'fingers of instantiation'. Most other authors I cite (Green, Quilty-Dunn, Kahneman, etc.) write of 'object files'. The two are slightly different theoretical posits; a FINST is just whatever is required to keep track of a visual location, an object file is more like a file for keeping information *at* that location. Both theories are founded on very similar empirical data, and the real difference is in the phenomena that data is used to explain. I will write of 'object files', but in light of this I take Pylyshyn and his collaborators' work to support mine.

seems unlikely that there are gustatory or olfactory objects in the same way.[57] We should not thus assume that what goes for vision will go for the other senses.

In the context of this thesis however, this is more of a helpful boundary than a worrying objection. I am not concerned with the nature of perception as a whole, but with the nature of those aspects or elements of the perceptual system which bear content. Any element of any of the senses which can bear content of this kind surely *will* relate to objects and their properties, and thus the work on vision will be generalisable to them. Anything else will be beyond such generalisation, but for just that reason it will be beyond the scope of this thesis, so this is hardly a problem.

#### 3.3.3.1 Against experiences

It is widely assumed that if there is any sense in talking of 'perceptual content', then it is perceptual experiences (paradigmatically visual experiences) which *have* this content. For example, S. Siegel (2005) writes:

> In contemporary philosophy, the phrase 'the contents of perception' means, roughly, what is conveyed to the subject by her *perceptual experience*. For example, suppose you are looking into a piano at the array of hammers and strings. There will be a way these things look to you when you see them: they will look to have a certain shape, color, [sic] texture, and arrangement relative to one another, among other things. Your *visual experience* conveys to you that the piano has these features.
>
> (my emphasis)

Notice that, though my perceiving is only described in terms of *one* object which I see (the piano) and its features, it is assumed that this is best understood as being simply the contents of my whole experience, not (for example) of some *part* or *aspect* of my experience which relates only to the piano leaving other parts or aspects to relate to other things.

Many of those I cited on perception in the first chapter make similar assumptions: e.g. Millar (2015 '… according to which perceptual *experiences* are …'), Schellenberg (2010 'There are two radically different conceptions of perceptual *experience*'), Soteriou (2000 'In discussions of the connection between the particularity of a successful perceptual *experience* and the intentional content of *experience*'), etc. (all emphasis mine).

[57] See Strawson (1964 chapter 2), Evans (1985 a response to Strawson), O'Callaghan (2008), and Nudds (2009) on objects in hearing; see Batty (2010) on olfactory objects. I have been unable to discover work on the possibility of haptic or gustatory *objects* in particular.

Of course, many of these authors may simply be operating with notions of content importantly different to mine. But even if it that were true, many of their notions must contain, entail or include mine. Finally, even if some do not, the main point here is that it is widely assumed that *any* notion of perceptual content must attach to whole perceptual *experiences*. Other things equal, this claim is often imported into work on perceptual content as I understand it. I will argue that this is a mistake in my project (though to be clear, this is compatible with it being entirely correct with respect to other projects).

The problem is that the notion of content I am working with is one which relates (or appears to relate) perception to *objects* (i.e. failure cases, Frege's puzzle, alternative- and non-satisfaction are spelled out in these terms) but in almost all cases, *whole experiences* are not what is related to objects.

The dialectic here is similar to that of §3.3.2.2. In that section, I argued that views of cognitive refemes as whole cognitive episodes (or their whole contents) were wrong, because such episodes can relate to more than object, but refemes each relate to at most *one* object. My argument here is similar: to make sense of the puzzles in the perceptual domain, we must think of perceptual refemes as each relating to at most *one* object. But perceptual experiences *don't* (always) each relate to one object. So perceptual experiences can't be perceptual refemes.

The piano case above is a good example: Siegel is concerned with whether or how well the *piano* is perceived. Surely if one perceives a piano under normal conditions one will perceive many other things apart from the instrument: the piano stool, the carpet, the wall behind the piano, etc. (one will also perceive many different things *in* perceiving the instrument: each key, the lid, the hammers and strings, etc.).

The piano case is very typical. Most normal day-to-day perceptual experiences are of this kind. Often what is perceived[58] in an *experience* is not just a single object, in terms of which Frege's puzzle, or an alternative-satisfaction case could be formulated. Contraposing, it is not always the case (though it may be sometimes) that a perceptual experience is of a single object. Instead, what is perceived in an experience is often a whole *collection* of objects. This point applies equally well to entirely non-veridical experiences: a hallucination of a piano and a stool is still a hallucination of two objects. Of course, some experiences may be of only one thing, the point is just that it is not part of the nature of perceptual experience that only one thing be apparently perceived.

This plurality of the objects of whole experiences means that the four cases we began with cannot (all) be described in terms of whole experiences.

[58]Or according to one's theory: presented directly, presented indirectly, represented, etc.

As before, consider alternative-satisfaction cases. In such cases there is a perceptual refeme which in fact relates uniquely to one object, and is associated with some property or description, which is satisfied by something *other* than that object. Consider the following case, modelled on Grice's (1961) two pillars. I seem to see to a cup of coffee in front of me. Unknown to me, the cup (and the coffee) causing this are off to the right, out of my field of vision, and reflected in an angled mirror which I do not notice. Behind the angled mirror is a qualitatively identical cup of coffee, exactly where it would have to be to affect me as the cup in the mirror *is* affecting me, were the mirror not there. Beside the mirror, but not reflected in it, is a jug of milk which I also see (this is a very thin description of this scene—in reality there would probably be more objects too).

Grice's case is supposed to be an alternative-satisfaction case. But this experience of the cup and the jug is not describable in those terms. Alternative-satisfaction cases require that the refeme relate to one object (and be associated with some property which is supposed to pick out that object uniquely but in fact does not). Which object does this experience relate to, with respect to which an alternative satisfaction case could be constructed? Clearly the question is misplaced as it relates to at least *two* objects: the cup and the jug. So whole perceptual experiences of objects simply cannot support description in the kind of terms necessary to be refemes.[59]

Similar reasoning would follow for non-satisfaction cases. Consider a variant on the above case: there is a bright red ball off to the right, out of my visual field, and there is an angled magnifying mirror with a colour filter in front of me. I do not notice the mirror, but I do notice the reflection of the ball (of course, I do not think it is a reflection, I seem to just see a ball). The overall effect is as if there were a light pink ball, much larger than the actual ball, directly in front of me. Finally, there is dark blue book slightly to the left, which I see as normal. This is clearly a non-satisfaction case: there is clearly a ball I *see*, however poorly, and this experience presents that ball as being a certain way (large, pink, etc.). But in fact that ball is not that way; were I to try to pick out the thing I see by locating whatever had the properties it seems to have, I would not locate anything, because *nothing has those properties*. Indeed, *nothing* is that way—there is nothing which satisfies the satisfiable content of the experience, but nonetheless I do see a certain thing.

Again, the *whole experience* is not describable in these terms. Some aspects of my experience constitute a non-satisfaction case (those related to the ball), and

[59]Notice again the clear parallel with §3.3.2.2—this objection is structurally just like pointing out that propositions do not have singular referents. This parallel would be a straightforward inference for those like Searle (1983) who think the contents of both whole perceptual experiences and whole cognitive episodes are propositional.

other aspects do not (those related to the book). Non-satisfaction cases require that the refeme relate to one object and be associated with some property which is supposed to pick out that object but in fact does not. Which object could the *whole experience* be said to relate to, with respect to which a non-satisfaction case could thus be constructed? Clearly the question is misplaced—it relates to at least *two* objects: the ball and the book.[60] As promised then, it emerges that perceptual experiences cannot play the relevant kind of role to bear content—perceptual refemes cannot be experiences.

Finally, a similar argument applies to hallucinations. In paradigmatic cases of hallucination, it is not the case that my whole sensory experience is non-veridical. Some parts are, some parts are not. Suppose I hallucinate a duck on the desk in front of me, and otherwise see normally. Everything I seem to see apart from the duck is real, so a great deal of my visual field is taken up by things I see. Nevertheless this is clearly a hallucination. It would be a mistake to insist that my visual experience *as a whole* is either veridical or non-veridical. Rather, some component or aspect of it is non-veridical—the one which seems to present a duck—and the rest is veridical.

In this, my discussion of perceptual content follows exactly the same pattern as §3.3.2.2. For both domains, I presented considerations like these to show that large complexes (whole thought episodes, whole perceptual experiences), related to *many* objects could not be refemes. This was a negative argument, now let us turn to a positive one.

#### 3.3.3.2 For files

So much for experiences. In both of the previous sections on linguistic and cognitive refemes, I argued against certain views of what refemes might be (that they were concepts, or propositions, or names, etc.) and the contrasting positive view of what refemes *are* was fairly clear. It is much less clear here; what could perceptual refemes be if not experiences?

Clearly, perceptual refemes must be some element of the total system of perceptual experience which is capable of relating to external objects, relating to at most one such object at a time, and being associated with certain properties which that object is supposed to have (in some cases, regardless of whether that

[60]Soteriou (2000) discusses a very similar non-satisfaction case, described in terms of whole experiences. It is telling that Sainsbury (2005 pp. 249–51) finds it necessary to redescribe this case in terms of the experience's veridicality *with respect to certain objects*. I have avoided Soteriou's example because it relies on 'displacing glasses' which displace the apparent location of one object, but not the whole scene, and it is not entirely clear to me how this is to be understood.

object actually has those properties). There is wide support in cognitive science for an entity of exactly this kind. They are called *object files*.[61]

An object file is an element of the mental and/or perceptual system, whose function is to track exactly one persisting external object, and store some information about this object. The relation between file and object is generally assumed to be *reference* (e.g. Echeverri 2015; Green 2018; Green and Quilty-Dunn 2021).

Object files are dependant upon the apparent presence of their referent—no apparent thing, no file.[62] Of course, there may or may not actually *be* a referent, since object files can occur in hallucination. The point is just that there cannot be an object file without the subject seeming to perceive some *thing*.

Object files are spatially-arranged (or spatially-indexed)—we use object files to keep track of the *relative* positions of different objects we perceive. Finally, object files store some kind of predicative or ascriptive information on their referents. The vast majority of evidence for object files has come from vision science, though I see no reason to believe they could not play a role in non-visual perception of objects, if such perception occurs.

I will not rehearse the empirical arguments for object files at any length (Green and Quilty-Dunn 2021 provide an excellent overview). Suffice it say that the evidence is significant, widely replicated, and (as far as I am aware) universally accepted; there are disagreements over exactly how object files function and what their content is, but the jury is very much *in* on whether there are any object files at all.

Much of the experimental evidence requires subjects to track one or more objects through the visual field as they move independently of one another. So during a single perceptual experience, there may be more than one object file active, each potentially relating to a different object.

It should thus be clear how an account of perceptual refemes as object files avoids the problems of an account on which they are experiences. The problem for the experiences-based account was this: to correctly describe alternative-satisfaction and non-satisfaction cases, refemes must relate to at most one object. But perceptual experiences are often of many more than one object at once, so cases which have been treated as clear paradigms of alternative-satisfaction are not properly describable as such if perceptual refemes are taken to be experiences.

If perceptual refemes are object files though, this problem is dissolved. Consider again the coffee cup and the jug: my *whole experience* (of the whole scene)

[61]This term was first used by Kahneman and Treisman (1984), and later popularised by Kahneman, Treisman, and Gibbs (1992).

[62]In this, object files are to be sharply distinguished from *recognitional* (ibid.) or *mental files* (as in e.g. Recanati 2012) which persist between occasions of activation.

does not constitute an alternative satisfaction case. Rather, my cup-referring object file does: my file in fact relates to the cup off to my right, but the file is spatially indexed and associated with visual information which make it appear that the thing I see (which is in fact the cup off to the right) is right in front of me. Moreover, *there is* a cup, exactly where it seems to me that the cup I see is, and with qualities which would appear to me in exactly the same way. So it is true that, were the mirror not there, I would have produced an object file indexed to the same location and associated with the same information (and thus indistinguishable from the object file I have actually produced), but which *in fact related to a different object*. This is a paradigm alternative satisfaction case. But this is not true of my object file for the jug, since the jug is exactly where it appears to me to be (unlike the cup). So, the jug does *not* constitute an alternative satisfaction case. By taking the bearers of perceptual content to be object files, the confusion has been entirely resolved.

The case of the ball and the book can be redescribed in a similar way. I have two object files: one relates to the red ball, and the other to the blue book. The former constitutes a non-satisfaction case and the latter does not. When I hallucinate a duck, I have a non-referring object file which seems to me to refer to a duck and some referring object files for other objects. The former is a failure case, the latter are not.

An important clarification is due here. Many of those who discuss object files in a philosophical context call them *representations*, and might thus be taken to hold both:

1. A roughly representationalist view of perceptual contents, i.e. the view that perceptual content is a representation of worldly objects (not, for example, the objects themselves, as a naïve realist would have it).
2. A view of the metaphysics of representations in such a representationalist view, namely that the representations are object files.

My argument is for neither 1 nor 2. Rather, my argument is for the claim that, if there is any sense in talking about perceptual content in an effort to solve the puzzles I began with, then we should take it that what *has* this content is object files (and not perceptual experiences). This is compatible with pretty much any view of what the content *is* (as I said earlier, that question is not the topic of this chapter), so this chapter is neutral on 1. 2 presupposes 1, so my claim is *a fortiori* neutral on 2 as well, though the claim I *do* make (that the bearers of perceptual content are object files) coupled with assent to 1, would entail 2.

## 3.4 Conclusion

Let us take stock. The purpose of this chapter was to describe the metaphysics of refemes in three different domains: linguistic, cognitive and perceptual. This goal was significant for two reasons: first, it was not independently clear what the metaphysics of refemes in any of these domains are, but clarity will be important for generating examples and test cases in later chapters. Second, various good candidates for refemes across these domains have wildly different metaphysics, so we might worry whether they can really have the same kind of content, or affect each others' content in the ways often supposed. We are now in a position to evaluate our goal in terms of these reasons.

I have argued that linguistic refemes are uses of terms (§3.3.1), cognitive refemes are mental tokens of individual concepts or similar structures, such as mental files (§3.3.2) and perceptual refemes are object files (§3.3.3). This constitutes an answer to the initial question, and should also make it clear how to properly articulate examples. Moreover, my method of arguing for these claims has been to argue for certain interpretations of various canonical cases from the literature, so these cases—*interpreted in this way*—will also serve as good examples later on. Thus the first reason for wanting good metaphysical definitions has been satisfied.

The second reason has been satisfied too. There were three kinds of apparent differences between the kinds of thing we might reasonably think are refemes.

First, there were metaphysical differences: names are long-lived entities, tokenable by different speakers at different times, but (on some models at least) thoughts and perceptual experiences are short-lived episodes unique to a thinker or perceiver. The worry was that these were good candidates for linguistic, cognitive and perceptual refemes, but that these things are so metaphysically different we should, *prima facie* be skeptical of their bearing the same kind of content.

This worry turns out be unfounded, because in fact all refemes are quite similar; utterance tokens, concept/file tokens and object files are all metaphysically short-lived entities, unique to an agent who tokens them. While there are obvious metaphysical differences (utterances are not object files are not cognitive tokens) they do not seem to me the *kinds* of differences which should make us skeptical that these things can bear the same sort of content.

Second, there were differences in the kind of content borne. I said above that one can have a thought or perceptual experience of more than object at once, but names (or their utterances) are generally thought to refer to at most one thing at a time.

This is entirely true—names (or their utterances) *do* differ from thoughts and perceptual experiences. But we have seen that thoughts and perceptual experiences are not refemes (neither are names for that matter). The comparison for my purposes is between utterances of terms, tokens of concepts or mental files, and object files. And we have seen that these things do *not* differ in this respect. Indeed, it was an important part of my argument that refemes can relate to at most one object, so things like perceptual experiences cannot be refemes. The disanalogy between the kinds of content borne is thus also not worrying.

I also said that *which object(s)* a single perceptual experience concerns might change over time, though the same does not seem true of thoughts; if I hold the same thought in mind for 60 seconds, it does not seem that I can be thinking one thing at the beginning and another at the end, whereas I might have a continuous perceptual experience as of one thing passing away and then a different one arriving. The same is probably true of names (recall 'Madagascar') but not utterances.

This problem, at least in this form, also dissolves if we take perceptual refemes to be object files. The original concern was that a single, continuous perceptual experience might involve conscious perception of two different objects (over the course of the one experience), but that if a subject thinks of one object then thinks of another there have been *two* thoughts instead of one. In other words: perceptual experiences seem to be individuated by some sort of continuous consciousness, whereas thoughts are individuated at least in part by which objects they are about. If thoughts and perceptual experiences are refemes, then there is a worrying discrepancy in the way their content functions.

This discrepancy does not arise for refemes as I have described them. Object files do seem to be partly individuated by the objects they relate to: if I have a single perceptual experience of one object passing away and another arriving, there are *two* object files at play overall. Clearly tokens of individual concepts or mental files, and utterances of terms also relate to at most one thing, so there is no worrying discrepancy.

A similar problem does arise in a slightly different form, to which I will return later. In cases of optical illusion, a single object file might track what appears to be a single object, but is in fact more than one. Consider for example a card trick where a magician waves a king in the air and during the motion, it 'changes' into an ace. There are actually two real objects here—the king and the ace—but the trick works because it appears that there is only one. I call these *transformation cases*, and will return to them in the following chapters.

Finally, there were differences in how different kinds of refemes might have their content fixed, especially if one thinks (as many philosophers do) that some linguistic refemes have their content fixed by their relation to other refemes. The discussion in this chapter has no bearing on this concern, but the issue of how content is fixed will be dealt with at length in Chapter 5, where I return to it.

# Chapter 4

# The Nature of Content

This chapter describes my theory of what content *is*. The theory is very simple. Some refemes have a referent; they refer to things. All refemes have ascriptions; they ascribe things to their putative referent. These ascriptions are the kind of thing which can be *satisfied*—in particular, by the referent (if any) of the refeme whose content they are part of. Content is just the combination of referent (if any) and ascription.

All refemes have content, since all have some ascription, and some have a referent as well. Some refemes (not all) have both. So we might say that content comes in two exclusive kinds (referential content on one hand, ascriptive on the other), or that it comes in two overlapping kinds (*merely* ascriptive, referential-and-ascriptive), or finally that there is one kind of content that has two aspects; these are all different ways of putting the same basic idea and I will use these different kinds of language interchangeably. When all is well and refeme-producing agents function as we should, a refeme has a referent which satisfies the ascriptions, though things can go wrong in various ways.

A structural analogy: consider a grading system in which every student who hands in receives written feedback on their work and all but the very worst pieces of work also receive a letter grade. So every piece of work (refeme) will have written feedback (ascriptions), and some will have a letter grade (a referent). When things go as they should, the latter will correspond roughly with the former; better work, better feedback, better grade.

There are three important structural features here. First, getting written feedback is sufficient for getting some feedback. Even the students who do not receive a letter grade at least learn something about their work. Second and conversely, getting both written feedback and a letter grade is different from getting only written feedback. Contrast with a system where all students get an achievement grade and an industry grade, both from one to five, which are then aggregated; in this system getting one and four is just the same as getting

three and two, but the system with written feedback and letter grade is not like this. Third, when all goes well, the two kinds of feedback match: students with more positive written evaluations will receive higher letter grades. Things don't always work out though—biased or overworked teachers give can give the wrong marks and curved grade policies can inflate or deflate letter grades compared with written feedback. Note this last feature is normative: there is a way things *should* be, but this is sometimes different from how things actually *are*.

On my theory, content is like the grading system: every refeme has some ascriptive content, and some also have a referent. Reference and ascription are separate things, but having either is sufficient for having content. Some refemes have both. When all is well, the referent and ascription are related in a roughly specifiable and predictable way—the ascriptions are satisfied by the referent—but this is not always actually the case.

These are the main claims of the theory, and they are substantiated and defended in §4.1. Everything else is detail. Most of this detail is an attempt to work out what must follow from the conjunction of the theory of content above and the theory of the nature of refemes in the previous chapter.

Most of the important details are about ascription. It is fairly clear what a refeme's referent *is*: some refemes refer and some do not. For those which do, their referent (the thing they refer to) is a constituent of their content (though contra views like Millianism, it does not *exhaust* their content). For those which do not the question does not arise, since they have no referent.

This leaves two significant questions about ascription, which I cover in §4.2. First, what is ascription? In §4.2.1 I claim that ascription is a set of properties judged by the agent producing the refeme to hold of its putative referent.

Second: how (if at all) are referent and ascription related? In some ways mine is a very old, very common kind of view. Those who have held views like it have also often held (or been thought to hold) that the ascriptive content generally or always *determines* the referential. This is, for example, the third thesis of descriptivism as Kripke understands it: 'If most, or a weighted most, of the [properties associated with a name] are satisfied by one unique object y, then y is the referent of [that name]'.[63] (1980 p. 71) Though Kripke isn't particularly faithful to the views of Russell, Frege or Searle, they did all believe something along these lines. I will call this claim Determination:

**Determination** *if content involves a satisfiable element (like ascription or sense) and an objective element (referent) then for any particular item of such content, the*

[63]Kripke writes as if some names are refemes. I do not believe this, but Kripke's articulating of the third descriptivist thesis can clearly be applied to refemes in general.

*objective element must simply be that, if anything, which satisfies the satisfiable element.*

Determination presents various problems for my view: for one, as I characterised ascription, it is possible for very many refemes to have the same ascription but different referents. For another, I have said only that a refeme's ascriptions *should* be fulfilled by its referent, but given Determination this will be true of any refeme with a referent.

In §4.2.2 I argue that Determination is false: what referent a refeme has is not, in general, determined by its ascriptive properties (i.e. is not in general just whatever happens to satisfy the ascription). This is an important departure from other theories. Even those who have held that content has two aspects, like Frege's sense and reference, have assumed the satisfiable element is in some sense prior to the objective one. In my theory, both elements are given equal weight, because neither is directly determined by the other.

This leaves us with a third question: what *does* determine which referent a refeme has? And a corresponding fourth, what determines its ascription? If we think of the question of the nature of content as something like a question of *semantics*, then these are analogously questions of *metasemantics*. These will be the topic of the next chapter.

With all this in hand, §4.3 describes solutions to the four puzzles with which we began. These should provide further motivation for the theory.

## 4.1 The structure of content

I argued in Chapter 2 that content is whatever solves certain kinds of puzzles. But then I argued in Chapter 3 that these puzzles must be interpreted in very specific ways if they are to have the dialectic force we want—in particular, for a certain view of what kinds of thing *have* content. So content (now in Chapter 4) must be whatever solves *those puzzles when they are interpreted that way*.

Taking on my view of refemes, notice that in order to properly describe a puzzle of any of the four kinds with which I began, it is necessary to attribute to each refeme both a referential status (either a particular referent or a stipulated *lack* of referent) and something non-referential but ascriptive. I argue that, because we *must* describe the cases in this way, these two aspects are a good model for content. The fact that the resulting model of content, given my comments in §4.2.2, *solves* those puzzles, is further, though separate, support for this view.

Let us take the four kinds of case in turn.

In failure cases, there is a refeme which does not relate to an object, but in some sense behaves as if it does—it is appropriate to say 'Vulcan was thought to be a planet' or even 'Vulcan is a planet' in a way it is not to say 'Vulcan is a balloon'. But such tokens of 'Vulcan' do not in fact refer. So to describe the case, we must include something *ascriptive* (being a planet, or being supposed to be a planet), but we must also stipulate that the refeme *does not refer* to anything. Leaving out either element is an underspecification: the case is interesting because 'Vulcan' refemes function as if they all refer to some certain thing which is a planet, *and* in fact do not refer to anything at all. To specify the case requires specifications of both reference-status and ascription.

Similar comments apply to perceptual failure cases: to explain a case of objective hallucination, one must say *both* that the subject seems to see something which is thus-and-so, *and* that there is in fact no object which they see.[64] The cognitive cases work in a similar way.

In Frege's puzzle, there are *four* refemes. In the linguistic version of the case, there is an identity claim including two tokens of the same term (call these U1 and U2 respectively) and then a second identity claim including one token of the same term again (U3), and a token of a different term (U4). So U1, U2 and U3 might all be utterances of 'Hesperus' and U4 an utterance of 'Phosphorous', thus:

**Trivial** *Hesperus is identical to Hesperus*

**Informative** *Hesperus is identical to Phosphorous*

To describe the case properly, we must then stipulate that U1, U2, U3 and U4 co-refer, and that Informative is *informative* in a way Trivial is not. Informativity has been understood in various ways, but a common idea is that U3 and U4 must differ in some way in which U1 and U2 do not, and which grounds the informativeness.[65] What way? Clearly not in what they refer to—they all refer to the same thing. Nor by the identity of the different refemes. For one, because U1 and U2 are just as metaphysically distinct as U3 and U4 are, for another, because this solution has the result that what we learn from Informative is about the refemes U3 and U4 (or at best, about the words 'Hesperus' and 'Phosphorous'), which seems wrong—we *learn* about Hesperus, not a token of 'Hesperus' or the word 'Hesperus' itself. It is rather, in every actual Frege case, some way that the

[64]Not just no object which *is* thus-and-so, but *no object at all*. Seeing an object which is really there, but seeing it 'wrong' is *misperception*, which is generally distinguished from hallucination.

[65]This rather minimal idea is even compatible with Campbell's (2002) logical notion of informativity, which I reject below.

agent takes the referent (if any) *to be*.[66] Once again, to describe the case properly, we must attribute to each refeme a referential status, and something beyond that mere status.

Non-satisfaction and alternative-satisfaction cases also require specifying both referent and ascription. In both kinds of case, there is something associated with a refeme which can be satisfied (for Kripke, the beliefs of the speaker; for Zarepour, the purely representational content of the viewer; for Soteriou, the properties *being round* and *being red*). But the cases are driven by a contrast between what (if anything) this satisfiable entity is uniquely satisfied by, and what the refeme *actually refers to*. Without both components, the cases cannot be properly described.

For example, Kripke's Feynman case (an alternative-satisfaction case) requires that there is a token which both refers to Feynman *and* is associated with the property *being a famous physicist* (the association relation here being *belief*). Without mentioning both components, the case cannot be specified in enough detail to do the philosophical work it is supposed to. Similarly with Godel/ Schmidt.

Another alternative-satisfaction case, Soteriou's (2000) round, red objects, is also like this: he has to say both that the visual experience the subject has is actually 'of' (i.e. refers to) one certain object, *and* that it appears to the subject to be a certain way.

Finally, the same is true of non-satisfaction cases. Recall Devitt's example of Jonah. Devitt describes how, even though the story of Jonah and the whale is almost certainly false (and probably not even substantial parts of it are true), it may be that modern uses of 'Jonah' do refer to someone who really did exist, around whom a legend has grown—we should say that Jonah did exist, and the story about him is false (or even stronger: that most of what we believe about him is false). Devitt takes this to demonstrate that whether an utterance refers (and if so, what it refers to) is not determined by the associated properties. To draw this conclusion we must acknowledge both that the modern utterances refer to a real person (objective part) and that there *are* properties which modern utterers believe to have held of this person (ascriptive part—though as it happens, these beliefs are mistaken). The case can't do the work Devitt wants without both parts. Similar remarks apply to Campbell's (2002) discussion of 'binding parameters'.

This is all to say: just in order to describe any instance of the four kinds of puzzle which originally motivated this dissertation, we must attribute to the refemes therein both a specific referential status (a referent or stipulated lack of referent) and some non-referential ascriptive content. Dropping either element

[66]This claim is to be read *de dicto*—i.e. it can be true even if there is no referent.

from a case renders it philosophically impotent; without it, the case would not be what it is. So, I claim, this two-factor structure is constitutive of these kinds of cases (with respect to their dialectical role). I argued in the first chapter that content is just that, an account of which will solve certain kinds of puzzles. If even *describing* the puzzles requires an attribution of referential and ascriptive status, then clearly content includes both. A refeme's content is the combination of its ascription and its referent (if any).

## 4.2 Ascription

It is fairly clear what reference is. A referent is a thing, and the relation of reference is some relation between refemes and such things. Putting aside the (many!) arguments over what exactly the relation *is*, there is at least some intuitive idea of the theoretical role that reference plays, and what it is like. It is much less clear what ascription is. This section is supposed to remedy that.

It may be helpful to begin by comparing the differing roles of the referential and ascriptive aspects of content.

The unified approach to the problems I urge in Chapter 3 is very agent-centric. I argue that we must understand the puzzles in terms of tokens produced at times, by agents (and not, for example, in terms of the timeless semantics of sentences or propositions). Within this kind of approach, talk of reference is used to describe constraints on what agents *should* do. Unless we mention which *actual* objects-in-the-world we are interested in, it is impossible to describe whether and how the agent *should* treat a token of 'Hesperus' and one of 'Phosphorous' in the same way or not. Similarly, unless we stipulate that a certain perceptual experience is a hallucination, we cannot say how the agent has gone *wrong* (even if they don't know it) by behaving as if they really perceive.

In this sense (and only in this sense) reference plays a normative role in our philosophical theorising.

My talk of ascription plays a different role: it is for describing what the agent *actually does*. One way to understand Frege's puzzle is a case where the agent behaves the *wrong* way (though without realising) because they treat tokens of 'Hesperus' and 'Phosphorous' differently when they should treat them similarly. The sense in which they treat them differently (i.e. what they *actually* do, as opposed to what they *should* do) is what ascription is supposed to explain.

Since reference and ascription play very different theoretical roles in the setup of the problems we are considering, it should be no surprise that it is necessary to stipulate both in a description of any such puzzle. At a high level then, the last three paragraphs just recapitulate §4.1.

Ascription also plays a satisfactional role. It is the ascriptive part of the content which makes alternative- and non-satisfaction cases what they are—by definition, ascription is something which can be *satisfied*.

### 4.2.1 The nature of ascription

So ascription plays two roles: it enables us to describe how an agent behaves (and whether they behave similarly or differently with respect to different refemes) and it must satisfiable.

Thus, I claim, ascription is a set of properties judged by the agent producing the refeme to hold of its putative referent. Talk of an agent's judgements is a good way to intelligibly explain what they do, in the way ascription is supposed to. And the properties can be satisfiable—here, in a specific way I describe below.

All refemes are produced by agents. When an agent produces a refeme, they make one or more associated judgements which can be expressed in natural language as an instance of the following schema:

**Ascription** *the referent of the token I am now producing is [is not, has, etc.] P*

where *P* is a variable ranging over terms for general properties or relations. These properties or relations *P* are the properties ascribed. Let us call these 'A-judgements'. In slightly more natural language, we might paraphrase all this into one of the following forms, as appropriate:

- the object I see/feel in *that* region of my visual/tactile field is [is not, has, has not, etc.] *P*
- the object I am thinking about is [is not, has, has not, etc.] *P*
- the object my present utterance of N [a name] refers to is [is not, has, has not, etc.] *P*

A-judgement is a special kind of judgement, namely one which can only occur with an associated refeme. An A-judgement associated with a refeme constitutively depends upon that refeme, and is 'about' the referent of that refeme, if any. Consequently, an agent's state of A-judging lasts only as long as the associated refeme exists.

The nature of the constitutive relationship between an A-judgement and the relevant refeme, and that between an A-judgement and the properties judged to hold of the referent, are both unanalysed primitives in my theory.

In the simplest case, for a refeme to ascribe some property is for that refeme to be associated with a judgement naturally expressible as an instance of Ascription with a term for the property replacing *P*. One sees an object which looks blue, tokens a visual refeme, and ascribes the property *being blue*.

This simple case only covers 'pure' properties like *being blue*, or *being a planet*—properties the having of which does not rely on the existence of a particular object.[67] The ascription of impure properties, like *being the same colour as IKB 79* (a blue painting) is also possible. We could think of the monadic-property A-judgements as having one 'slot' or 'place' for the object to which they are ascribing, though this is only a figurative way of talking. Relational-property A-judgements have more than one 'slot'. In impure ascriptions then, first the agent tokens the refeme whose content includes the ascription of an impure property, then this is associated with an A-judgement which ascribes a relation, which is in turn related to *other* refemes, for the various other 'slots'.

A worked example (more filled-out solutions to the original puzzles are below). Most people who have heard of Vulcan have two significant beliefs about it: that it was 'discovered' by Le Verrier (though in fact there was nothing to discover) and that it was supposed to be a planet. So on thinking of Vulcan, an agent tokens a refeme R1 which does not refer (since there is no Vulcan); they make an A-judgement that the property *being a planet* holds of the referent (if any) of R1; they also token a refeme R2, which refers to Le Verrier; and finally they make a second A-judgement that the referent of R1 is such that the referent of R2 believed he had discovered it. The content of the original refeme is a conjunction of: the property *being a planet* and the property *being discovered by Le Verrier* (R1 has no referent, so no referential content). Le Verrier enters the ascriptive content of R1 by combining an A-judgement with R1 and R2, because R2's content includes Le Verrier himself. Of course this is rather simplified—in reality each refeme probably has many more associated ascriptions than these. But this demonstrates the principle.[68]

A few more clarifications may also be useful.

First, A-judgements might be more explicit or more tacit. The difference is best brought out with normal, non-ascriptive judgements. Sitting down on a chair, I judge that the chair is there beneath me, though this state or process of judging is not something I am immediately aware of. Or referring to a feminine-looking person I have just met, I unthinkingly use pronouns 'she' and 'her'; I judge that she is a woman though without being aware of so judging. Coming to

[67]There are two slightly different notions of impurity. According to one, a property is impure if having it requires a relation to an object; thus being male is pure, but being a father is not. On this view, purity coincides with intrinsicality. On the other view, a property is only impure if having it requires a relation to some object in particular; being male is pure, and so is being a father, but being *Hugo's* father is impure. On this view there can be pure but extrinsic properties. The objection I consider is most forceful on this latter conception.

[68]Careful readers will notice that this example mentions *beliefs* as well as judgements, though belief has so far played no role in my system. The beliefs here are just meant to make it more intuitive why a case as described might actually occur (i.e. why an agent ascribes one property and not another). Exactly how this works will become clear in the next chapter.

the end of a difficult mathematical exercise, I judge that the answer is a certain number, and *this* state or process of judging I *am* highly aware of. In the first case my judgement is more *tacit*, and in the second more *explicit*. Much has been written on what this distinction really comes to, whether it is a binary or a gradable distinction, etc.. All I require is that there is some intuitive difference between how obvious, conscious or available the two judgements are.

Most A-judgements are more tacit. The agent making the judgement is not immediately aware *of* the judgement as they make it. On the model so far described, when I think about my friend Ed, who has brown hair, I token an Ed-referring refeme, and make an A-judgement that the referent of that refeme has brown hair. But in such cases I would very rarely be explicitly *aware* of making this judgement (or of tokening the original refeme).

Some A-judgements are explicit though. There is nothing in the nature of content, or of A-judgements, which requires that they be tacit, and philosophers thinking about such things are apt to go about making explicit A-judgements.

Second, I do not claim (and do not believe) that an A-judgement which specifies a refeme's ascriptive content is *part* of that refeme (or is identical with it, or constitutive of it). Certainly no utterances are constituted by judgements. The same is probably true of object files, and I am not concerned with the metaphysics of cognitive tokens. Rather my claim is that the properties judged to hold of the referent (partly) constitute the *content* of the refeme, just as the referent—if there is one—partly constitutes the content without itself being the refeme. The content of the refeme is a complex of properties, and sometimes an object, but never a judgement (unless of course, the referent of some refeme is an act of judging).

In fact, something close to the reverse is true, as I said above: A-judgements constitutively depend upon their refemes. A-judgements are a special kind of judgement which require, and depend upon, associated refemes, which fix what thing they are judgements about.

Third, I do not claim that A-judgements themselves have a structure which is, or is isomorphic to, the linguistic form I give above (as a proponent of the language of thought hypothesis might claim). The linguistic form is used *only* to express in writing what the judgements are like. I am not committed against this view either.

An important corollary is that my theory does not require that any content which concerns an object also concerns the token which refers to that object. This would be true, and would form the basis of an objection, on a simple language-of-thought reading of what I said above.

The objection would go like this: I say that when a normal person sees a book, they produce a perceptual token which refers to the book, and they make a judgement with the content *the referent of the token I have produced (or am producing) is a book*. But, the objector continues, this entails that the agent refer to the token they have produced, since the italicised clause representing the A-judgement includes the phrase '*the token I have produced*' (similarly the use of '*I*' entails that they refer to themselves).

This would be problematic. For one, it just doesn't seem very likely; the agent sees, and has content about, the book, not their own experience. For another, it places overly strong requirements on agents' ability to see, discuss and think about things, since many agents *cannot* entertain content concerning their own experience. Children and those with cognitive disabilities (including temporary ones, like being drunk) can see and recognise a book, though without having the higher cognitive function available to think *about* their act of seeing or thinking. So, the objector goes on, my theory requires higher cognitive functions of agents who do not (first case) or cannot (second case) exercise such functions. So my theory must be wrong. A similar objection could be raised with respect to self-reference and '*I*'.

Fortunately, this is a mere seeming. The objection infers from the fact that instances of the linguistic schema I have given include reference to tokens (which is true) to the fact that the content *which this schema represents* must also include such a reference. This would only be true if the relation between the representation and the content were a very strong isomorphism, as some believers in a language of thought would claim. But this is not something I am committed to; indeed, it is something I feel pushed to deny, for exactly the reasons the objector points out.

Moreover, the role that reference plays in the objector's reading is already played by a different relation, which I have already described, and which is not problematic in this way. The objection points to the fact that each A-judgement must be associated with a single refeme, and infers from the form of the linguistic schema that the relation of association is reference. But I said above that what relates an A-judgement uniquely to a refeme is *constitutive dependence*. Each A-judgement relates uniquely to the one refeme upon which it constitutively depends. Reference need not enter the picture.

Fourth, the phrase 'the referent' in the schema above is to be read *de dicto*, i.e. without the agent being committed to the existence of the referent or any of its properties by whatever feature of their judgement is represented by this phrase. This is why I sometimes write of ascription to a 'putative referent' above.

To make this clear, we might rephrase the form to the effect that the agent judges:

**Ascription** *the referent—whatever it is, if there is one—of the token I am now producing is [is not, has, etc.] P*

Though again, since I am not committed to the judgement's having a structured or sentence-like form, I am similarly uncommitted to the nature of the 'whatever it is' feature of the judgement, as it distinguishes judgements of this kind from others.

Fifth, this is not a theory of judgement, especially not *propositional* judgement. The theory given so far is one of how our utterances, thoughts and experiences relate to particulars (and associated phenomena, like how they *appear* but fail to so relate). I have not tried to give a theory of all mental or intentional phenomena, or of judgement. In particular, though I have talked a lot about objects, properties and our mental representations of them, I have *not* tried to analyse 'propositional' judgement.

The way I have written might suggest that that for an agent to judge that Hesperus is bright *just is* for them to token a Hesperus-referring refeme, which ascribes brightness. This view is entirely compatible with my theories, but it is not entailed by them; judgements like that might require more, and other kinds of judgement will probably require different analyses altogether.

Fifth, there are no constraints on what one agent might ascribe to the same actual referent. In particular, there are no logical, inferential or rational constraints; it is entirely possible for the same agent, at the same time, to ascribe P and not-P (or P and Q, where P and Q are inconsistent) to the same existing referent, or to the same merely-putative referent. Of course, it might not be ideal for them do so—it might be illogical, irrational or contradictory—but as far as my theory goes, it is entirely *possible*.

Finally, ascriptive content can change over the lifetime of a single refeme.

I wrote above that when an agent produces a refeme, they make one or more associated judgements about its properties, and these properties judged to hold of it are the ascriptive content. This is true of all refemes, and this model is sufficient to explain ascription in short-lived refemes like utterances of names; an utterance is over in an instant, and it is unlikely that the speaker would change their views or beliefs about the referent during their act of uttering.

But some refemes are much more long lived. Watching an object move across a room, an agent will token *and maintain* a single object file for that object, perhaps for several minutes. Or during a long discussion on a single topic, a

conversant may token and maintain a single cognitive token, relating it to various different utterances as they go along.

If the apparent properties of the perceived or discussed object change during this time, the agent will simply alter what properties they judge to hold of it. That is, they will *update* its ascriptive content. At every point, the refeme has some ascriptive content, though perhaps different content at different points. So a stricter version of my claim is that when an agent produces a refeme, and at every time during its existence, they hold one or more associated judgements about its properties, and these properties judged to hold of it are the ascriptive content.[69]

Of course, this stricter version will only be relevant to the kinds of refemes for which ascriptions can be updated. Since ascriptive content lasts only as long as the refeme which bears it, the content of a refemes with very short lives will never change. Utterances of names fall into this category. Other refemes, with longer lives, might have quite various content, as with the perceptual and cognitive refemes above.

### 4.2.2 Reference is not determined by ascription

It is widely thought that on *any* view of content (or anything suitably *like* content) which involves a satisfiable and an objective element, the objective element must simply be that, if anything, which satisfies the satisfiable element. I call this principle Determination. Frege is often read as holding this view. Pepp for example says:

> Frege's account of the mechanism of reference is that speakers, and the expressions they use, refer to things in virtue of those things being determined by the *Sinne* [sic] that they associate with their words. Frege suggests that to be determined by a Sinn is to be the unique satisfier of some condition.
>
> (Pepp 2012 p. 7)

Kripke (1980) takes the 'Frege-Russell view' to claim that 'If most, or a weighted most, of the [properties associated with a name] are satisfied by one unique object y, then y is the referent of [that name]'. (p. 71) A similar reading of

[69]This version of the claim deals with judgements 'held' over time, rather than (and as well as) those 'made' *at a time*. These might be better understood as beliefs, since they are more like standing states than momentary actions; a belief-based way of talking could even be extended to momentary judgements with which I originally framed ascription. I don't think this distinction makes any substantive difference to my overall views, and it would introduce unnecessary complexity. In the rest of the thesis I just write of 'judgements', though I am sure my claims could be reinterpreted if necessary.

Frege is given approvingly by Campbell (2002), of which more below, and Crane writes 'Some think that the content of an intentional state must determine its object'. (2009 p. 477)

Determination also shows up in the literature on objective perception. Schellenberg (2016) discusses a view she calls 'generalism' on which 'perceptual states are constituted only by general elements'. She says:

> Generalism is attractive for a number of reasons [but these come at the cost that it] does not give an adequate explanation of how it is that we single out particulars in perception by discriminating them from other particulars, thereby yielding perceptual knowledge of particulars, justifying singular thoughts, grounding demonstrative reference, fixing the reference of singular terms, and providing the basis for singular thoughts.
>
> (Schellenberg 2016 p. 27)

It is true that generalism alone does not offer an account of these things. But why should it? Surely we would only expect an account of the contents of perception to furnish an account of perceptual individuation or reference if we assume that the former determines the latter (i.e. that on a generalist picture, the general contents of perception determine perceptual reference). So Schellenberg assumes Determination (this assumption is especially stark in Footnote 39).

G. Siegel's (2022 p. 2387) reading of Pylyshyn's comments on object files are similar. He says: 'The referents of the object file represensations are plausibly determined by an object's estimated spatiotemporal location'. If one takes spatial location to be the kind of property which might be encoded in an object file, this amounts to the claim that Determination is plausible. Echeverri (2015) also discusses this view with respect to perception (see the 'satisfactional claim' p. 286).

Determination is often used, together with alternative- and non-satisfaction cases, to argue against views of content like the one I suggest. For if Determination is true (or at least if the view entails it), then those cases are counterexamples to the view. The dialectic goes like this. There is a certain view of content (semantics, perceptual content, etc.) according to which:

1. The semantics of some meaning-bearer involve both its reference (≈ denotation) and
2. something like a description, predication, ascription or sense.
3. The latter determine the former.
4. Where 'determine' is understood as unique satisfaction, so that the referent is whatever uniquely satisfies the ascription.

3 and 4 together constitute Determination. This is an attractive view, and Russell and Frege both held something like it. Alternative- and non-satisfaction cases—most famously Kripke's Gödel/Schmidt case—demonstrate conclusively that the view as a whole is false.

The popular Millian reading of Kripke takes this as a signal to reject the view wholesale (or at the very least, to retreat to a theory which only includes 1).

Sadly for this reading, 1 and 2 are clearly basically right. The idea that content is a conjunction of a *thing* and the *way that thing is presented* has persisted for a reason.

Happily, Neo-Kripkean Millianism is not the only response. Evans (1982) could easily be read as an attempt to rescue the general picture. He retains 1, 2, and 3 (he says 'information' instead of 'ascription'), but rejects 4, saying that what *causes the information* is the referent, regardless of whether that satisfies the information's 'content'. I see myself as engaged in a similar project to Evans, but I reject 3 and 4 entirely; that is, I reject Determination. My account is completed by an alternative picture of what determines reference, as well as one of what determines ascription, in the next chapter.

I take the success of my own picture, and the clear falsity of 1–4, as positive reason to reject Determination. I am only aware of two positive arguments for Determination; in the rest of this section I argue against each in turn.

#### 4.2.2.1 Campbell on Frege and Informativity

The most sophisticated argument I know of for Determination in any context is Campbell's (2002). Campbell is discussing Frege's view that the sense of a name determines its reference. Clearly the theories are different, but I take it that there is a close enough analogy between Frege's view and mine that a fruitful comparison can be made. Frege takes refemes to be things like terms, tokenable on more than one occasion, but themselves independent of those tokens. Frege thinks that each refeme has a sense and a referent, the sense is a mode of presentation of the referent, and that there is some way in which sense determines reference. This last claim has been variously interpreted. Campbell claims:

> It does follow from the characterization [sic] I have given so far that sense must determine reference, in that sameness of sense guarantees sameness of reference. For suppose we had sameness of sense of two terms, but not sameness of reference. Then the identity statement connecting the two terms would be uninformative, or a logical truth,

> because of the sameness of sense, yet false, because of the difference in reference. And that hardly seems possible.
>
> (Campbell 2002 p. 85)

This in turn relies on a particular notion of informativity:

> This emphasis [in Frege's description of his puzzle] on the notion of informativeness is unfortunate, because it has sometimes led commentators to think it telling to ask the question, 'informative to whom?' After all, you might say, if I already know that the Morning Star is the Evening Star, then it will not be informative to me to be told that the Morning Star is the Evening Star; whereas if I did not know that already, the remark will be informative. This can make it seem that Frege's criterion makes sameness or difference of sense depend on the idiosyncrasies of your prior state of knowledge. I think the right reaction to this comment is to put Frege's criterion directly in terms of what is, and what is not, an instance of a logical law. To be 'uninformative', in this sense, is to be an instance of a logical law.
>
> (Campbell 2002 pp. 84–5)

Throughout the rest of the chapter, Campbell uses the term 'logical truth' interchangeably with 'instance of a logical law'.[70] Thus a more explicit reconstruction of Campbell's argument runs:

1. Two terms have different senses iff an identity claim made between them is informative (= is not uninformative).
2. All uninformative statements are logical truths.
3. Suppose a,b have the same sense and different referents.
4. The statement 'a=b' is uninformative. (1,3)
5. The statement 'a=b' is an instance of a logical truth. (2,4)
6. The statement 'a=b' is true. (5, all logical truths are true!)
7. The statement 'a=b' is false. (3, difference in referents)
8. The statement 'a=b' is both false and true. (6,7)
9. There are no a,b with the same sense but different reference. (7, RAA and reject supposition 3)

[70]It has been suggested to me that 'logical law(s)' in Campbell's work might be most charitably interpreted as the law of identity, such that all uninformative utterances/statements are identity utterances/statements. Since all my and Campbell's examples are of identity claims, the following argumentation works just as well on this interpretation as on a more general one (i.e. concerning *all* logical truths). However, I think the more general interpretation is closer to what Campbell himself had in mind, so it is the one I follow in the text.

Frege (and Campbell) are working in terms of *terms* (like names) having sense. My refemes are not terms, they are utterances or similar event-like tokens. Frege's and Campbell's ideas are applicable to these too though, and the argument can be easily reconstructed with utterances:

1. Two utterances have different senses iff the identity claim made using them is informative (= is not uninformative).
2. All uninformative utterances are instances of (sentences expressing) logical truths.
3. Suppose a,b have the same sense and different referents.
4. The utterance 'a=b' is uninformative. (1,3)
5. The utterance 'a=b' is an instance of a logical truth. (2,4)
6. The utterance 'a=b' is true. (5, all logical truths are true!)
7. The utterance 'a=b' is false. (3, difference in references)
8. The utterance 'a=b' is both false and true. (6,7)
9. There are no a,b with the same sense but different reference. (7, RAA and reject supposition 3)

If there are no refemes a, b with the same sense but differing reference, then it is true that sameness of sense guarantees sameness of reference. This was Campbell's gloss on Determination, so the argument seems to show that Determination holds.

My response is to reject 2. Without 2, the inference from 4 to 5 is invalid, so we cannot obtain 6, so there is no (guaranteed) contradiction.

Campbell thinks 2 is true because if it isn't, then whether a sentence is informative may depend on one's knowledge state, and in turn the sense of the names in it will also depend on this (because sense is just whatever a difference in which makes for informative identify statements). Campbell doesn't elaborate on why this is problematic but we might have at least two worries (which I presume Campbell would share). First, who's knowledge state could it depend on? The speaker's? The hearer's? Terms, including names, are used by many different speakers and hearers throughout their lives, so which agent's state should have the privilege of determining the reference seems arbitrary at best. Second, knowledge states vary with circumstances, so whether two terms have the same sense will change depending on the circumstances.[71]

These are valid concerns with regard to the sense of *terms* (so perhaps Campbell's original argument holds water). But we have seen already that linguistic refemes are utterances, not terms. Thus the two reasons are not so worrying. First, for each utterance it is always clear whose state is in question. In nearly

[71]In particular, on whether certain agents know they co-refer, and not on circumstances which we might expect to alter a term's sense, like a significant change in patterns of its usage.

all cases, it the producer's. Most utterances of names are produced by a single speaker, speaking normally to express their own beliefs or wishes. In these cases, since they are the sole producer, only their knowledge is in question. Utterances might also be produced by groups: imagine a group singing Happy Birthday and saying a name in unison, or each saying a single word to compose an extended sentence (whether planned or not). Much of the literature on the nature of group utterances recruits similar resources to that on group intentionality and knowledge (see e.g. citations in Ludwig 2020); the conditions for a group utterance occurring when many agents speak together are similar to those for group intentionality or knowledge when many agents intend or know together. So in these cases too, the agent whose knowledge is in question will be the same as the producer: it is the relevant group.[72] Finally, an utterance might be produced by one agent on behalf of another, when the former acts as a spokesperson (Hughes 1984; Ludwig 2013; 2020; Lackey 2017 all make much of this). In these cases, the relevant knowledge state is that of the latter agent (on whose behalf the spokesperson speaks), not of the spokesperson themselves.[73] Nonetheless, there is still clearly a *single* agent, upon whose knowledge state the features of the utterance would rest; it just so happens that the utterances are actually articulated by someone else.

The second problem is an artefact of Campbell's (Frege's) preference for talking about *terms*, where a term is something which can be deployed in different circumstances by different speakers. But utterances *are those deployments*—there is no sense in which one utterance can be tokened under two different circumstances. Each utterance is fixed to a particular set of circumstances—the circumstances in which it is produced. The second problem simply does not arise, because contra Campbell, no refeme can occur in several different circumstances.

So the reasons I have put into Campbell's mouth for 2 don't hold, and we are free to reject it. This is no *reason to* reject it though—it just shows that the position is tenable.

[72]I will not commit here to whether there *are* group utterances or not. My claim is just that, if there are, then the knowledge those utterances' features depend upon is that of the group. This seems very likely, since much of the literature on the question of group utterances draws on resources from that on group agency and knowledge. Of course, if there are no group utterances, then cases of apparent group utterance will really just be several individual utterances, each treated separately as already discussed.

[73]As several authors point out, when the spokesperson is a member of the group whose knowledge they articulate, the spokesperson's individual knowledge may play some role in determining the group's knowledge, and therefore in what the spokesperson themselves expresses on the group's behalf. But playing this role does not mean that the spokesperson's knowledge *is* the group's knowledge; the spokesperson still articulates knowledge (the group's) which is not merely their own.

But we *should* reject 2. 2 says that 'all uninformative utterances are instances of (sentences expressing) *logical truths*'. For this to be true, it must at least be the case that every utterance is the type of thing that *could* be an instance of a logical truth. And that is true: utterances (of full, truth-apt sentences) are linguistic expressions, and so are exactly the kind of thing which can instantiate logical truths. Linguistic refemes will be subsidiary constituents of these utterances. Campbell uses this to argue for a version of Determination about linguistic refemes and their content.

However, I originally brought up Determination, and Campbell's argument for it, in my discussion of content for *all refemes*, including non-linguistic ones, and the real concern is whether it holds across the board. If Campbell's argument, and 2 in particular, is to convince us to hold Determination for kinds of content beyond just the linguistic, then the whole thing, including 2, must be transposed into a more general form. But as I will show, when read in this form, 2 is clearly false and so the argument fails.

The 'utterances' in my version of 2, or the 'statements' in Campbell's are complexes, which include refemes among their constituents. On my reading, the refemes of 2 are utterances of names; on Campbell's they will be the names themselves. Campbell's argument following 2 goes on to say things about these refemes. There are other kinds of complexes with constituent refemes: complex thoughts involve content-bearing cognitive tokens, and visual experiences include, but are not exhausted by, object files. A version of 2 concerned with content in general (in language, thought and perception) would have to be about these complexes in general. At a first pass, that might look like this:

2a. all uninformative propositionally-contentful complexes with constituent refemes are instances of (types of the same such complexes expressing) logical truths.

Campbell's argument would only prove Determination for non-linguistic refemes and content if 2a were true. But it is straightforwardly false. It applies to propositionally-contentful complexes in domains other than language (i.e. thoughts and perceptual experiences), and claims that some among them are instances of logical truths. But what would it be for a thought or an experience to instantiate a logical truth? Perhaps we can make sense of this for thoughts, especially if thoughts occur in some language-like structure. But surely perceptual experiences never instantiate logical truths. Logical truths exist within systems of symbols, with conventionalised or stipulated meanings. But there are no symbols or meanings in the perceptual system, so how could it instantiate

anything logical at all?[74] So on this first pass generalisation (2a), we should reject 2, and Campbell's argument for Determination fails.

There is also a reasonable second pass generalisation. It might be objected that 2a is a bad formulation. In moving from 2 to 2a, the term 'utterances' in 2 is replaced with the more general 'refemes' of 2a, but the 'logical truths' of 2 remains unchanged. It is precisely because the logical truths are not instantiable by all complexes containing all kinds of refemes that this formulation fails. Perhaps the interpretation in 2a fails because one element of 2 has been generalised, but not the other. In this vein, a more charitable interpretation would replace 'logical truths' with something else, generalised in step with 'refemes':

2b. all uninformative propositionally-contentful complexes with constituent refemes are instances of (types of the same such complexes expressing) Xs, where Xs are whatever in the relevant domain play the same role as logical truths do in language.

This interpretation fails too, and for similar reasons: whatever role logical truths play in language, it is one characterised by formal inference; since perception does not involve formal inference, nothing in perception can play a similar role. To be clear, perception may well involve reasoning, conceptual content and other things typically associated with inference (or it may not) but it certainly does not involve formalised relations between symbols with stipulated or conventionalised meanings—i.e. formal inference. So the second pass generalisation fails too.

Having rejected 2, Campbell's inference from 4 to 5 is unavailable, and thus we cannot derive Determination. Thus, Campbell's argument does not establish the general version of Determination of interest in this thesis.

#### 4.2.2.2 Searle on teaching names

Searle also argues for Determination:

> How, for example, do we learn and teach the use of proper names? This seems quite simple—we identify the object, and, assuming that our student understands the general conventions governing proper names, we explain that this word is the name of that object. But unless our student already knows another proper name of the object, we can only identify the object (the necessary preliminary to teaching

[74] A clarification: perceptual experiences do *follow* logical laws. For example, every perceptual experience is identical to itself and to nothing else (the law of identity) and for every putative fact P about perceptual experiences, it is true that $P \vee \neg P$ (law of non-contradiction). My claim is that no perceptual experience *instantiates* a logical law, in the way that an utterance of 'P or not P' instantiates the law of non-contradiction (but not the law of identity).

> the name) by ostension or description; and, in both cases, we identify the object in virtue of certain of its characteristics. So now it seems as if the rules for a proper name must somehow be logically tied to particular characteristics of the object in such a way that the name has a sense as well as a reference; indeed, it seems it could not have a reference unless it did have a sense, for how, unless the name has a sense, is it to be correlated with the object?
>
> (Searle 1958 p. 168)

Searle seems to assume throughout that the rules for using a name constitute (partly or fully) the meaning of that name. I take Searle's argument to go like this:

1. To correctly learn or teach the meaning of a proper name *N* we must identify the object *N* names, but without using *N*.
2. Often (perhaps ultimately) we must identify the object *N* names without using any names at all.
3. Instead of using a name, we must identify the object *N* names by demonstration or description.
4. Identification by either demonstration or description picks out the object 'in virtue of certain of its characteristics'.
5. For at least some proper names *N*, in order to learn or teach the meaning of *N* one must use particular characteristics of the object.
6. These characteristics constitute rules for using the name, in particular, they constitute rules for *which* object the name is correctly applied to.
7. The object it is correctly applied to is its referent
8. So, if the name is associated with characteristics (which Searle thinks it is) its referent (if any) *must* be that which satisfies those characteristics.

In other words: if names have anything like ascriptive content associated with them, then that content is whatever is used in correctly teaching the use of the name, and in order to be used in *correctly* teaching the use of the name, it must be uniquely satisfied by the referent, because picking-out-by-unique-satisfaction is part of how the teaching process works. So if a two-aspect theory of content like mine is right (for names at least), then the ascriptive part determines the referential part.

I am fairly sympathetic to this view regarding *names*. But this thesis is not about names, it is about refemes, and linguistic refemes are *utterances*, not names. Searle's argument is only of interest in so far as it can also be made with respect to utterances. Unfortunately, it cannot. Consider each step in turn:

1. to correctly learn or teach the meaning of an utterance *U* of a proper name *N* we must identify the object *U* refers to, but without using *U*.

This is true. Utterances of proper names have meanings, and to explain these meanings to those who don't know them, we have to identify the object referred to. Since the student doesn't know what the utterance means, we can't rely on the utterance itself to do this. If I say to my friend at a party 'I think Robin is here' and my friend doesn't know who Robin is, it will not help to explain that they are the referent of my last utterance of the name 'Robin'.

2. Often (perhaps ultimately) we must identify the object *U* refers to without using any names at all.
3. Instead we will have demonstrate or describe the object.
4. Identification by either demonstration or description picks out the object 'in virtue of certain of its characteristics'.
5. For at least some utterances *U* of proper names, in order to learn or teach the meaning of *U* one must use particular characteristics of the object the utterances refers to.

2–4 are also clearly true, and 5 follows as a consequence. To explain who I was talking about, I might say to my friend 'I meant the person in the corner, in the white shirt'.

It is not immediately clear how to transpose 6 from Searle's talk of names to my talk of utterances. Clearly the rules must concern utterances beyond the one originally under discussion—a rule for interpreting a single utterance is not terribly useful. The rules could be taken to be rules for producing utterances:

6a. These characteristics constitute rules for making further utterances of the same name. In particular, they constitute rules for *which* object such utterances are correctly applied to.

Or for interpreting utterances:

6b. These characteristics constitute rules for interpreting utterances of the same name (including the original utterance). In particular, they constitute rules for *which* object such utterances are correctly taken to apply to.

Or both:

6c. These characteristics constitute rules for making and interpreting utterances of the same name (including the original utterance). In particular, they constitute rules for *which* object such utterances are correctly applied or taken to correctly apply to.

All three formulations are false. Suppose I explain to my friend that by my utterance of 'Robin' I meant the person in the corner, in the white shirt, and my friend thus comes to understand my previous utterance, and be in a position to make and interpret further utterances of 'Robin'. So the characteristics by which

I teach the meaning of my utterance are wearing-a-white-shirt, and standing-in-the-corner (and perhaps being-a-person). But it is clearly *not* the case that these constitute rules for the correct application or interpretation of utterances of the name 'Robin'. The person in question may soon fail to satisfy wearing-a-white-shirt and standing-in-the-corner, but later utterances of the same name would still be appropriately made in reference to them, and taken as referring to the them. So Searle's argument for Determination fails when extended to the kinds of thing which I have argued are refemes. (Though as above, this doesn't mean that Searle's argument could not succeed for *names*.)

## 4.3 Solutions to the puzzles

My theory of refemes and their referential/ascriptive content provides solutions to the four puzzles which originally motivated it.

The solutions given in this chapter are 'semantic' solutions: descriptions of what content the various refemes in each case have, in a way which usefully explains what is going on.

I said above that the two aspects of content play different roles in our theorising: reference has a normative role, and ascription has a descriptive one. When things go as they should, reference and ascription match up: on uttering 'Feynman' I think I am talking about a famous scientist, and indeed I am—the referent of my refeme has all the properties I ascribe.

The four puzzles with which we began are four ways things can go wrong and depart from this standard.

So one kind of solution to these puzzles is just to describe what has gone wrong: in what way cases of that puzzling kind are departures from the standard 'good case' where reference and ascription match. We might call these 'semantic' solutions: descriptions of the kind of content at play in each case. The solutions offered in this chapter are along these lines. What is more, they are structurally identical across language, perception and thought, as we should expect if Unity is true.

The solutions below might strike some as rather thin, or incomplete. Semantic solutions are to be distinguished from 'metasemantic'. Given the solutions of this chapter—useful descriptions of how each case departs from the 'good' standard—we might ask why such departures occur at all? What determines that the refemes have these defective contents rather than non-defective ones? Answers to these questions would fill out the semantic solutions, but they require a general view of the determination of content; all this will come in the next chapter.

### 4.3.1 Alternative- and non- satisfaction cases

These cases are the most straightforward. In an alternative-satisfaction case, there is a refeme with a clear referent, and clear satisfiable content, such that there is something other than the referent which satisfies the content. The problem is that the views these cases are raised against all include some version of Determination—if a token refers to something, that must be because the token is associated with properties which pick out that thing uniquely. If there are many satisfiers, and none uniquely so, the theory does not correctly predict which one is the referent.

In Kripke's 'Feynman' case, utterances of Feynman do refer to Feynman, even though all the speaker believes about Feynman is that he is a famous physicist, which is not enough to individuate him uniquely. This is only a problem if we expect the individuation of Feynman to proceed from the speaker's beliefs, but I have argued we should not, so the problem simply does not arise. Instead the correct analysis is that utterances of 'Feynman' do refer to Feynman, and that what the speaker generally believes or (at some time) judges about Feynman has very little to do with this; Determination is false. Similar remarks apply to Zarepour's Jones/John case, though the refemes are cognitive tokens, not utterances.

In Grice's clock case, the satisfiable content is something like *being a clock on the shelf in front of me*. And there is just such a clock on the shelf, but the case is setup so that this clock is not the referent. Again, this is only problematic if we assume that the thing, if any, which satisfies all the properties must be the referent. I have claimed we should not. Instead, the agent tokens an object file, and associates with it an A-judgement of being-on-the-shelf. But the actual referent, if any, of the object file is determined in a separate way (to be described in the next chapter).

In a non-satisfaction case, there is a refeme with a clear referent, and clear satisfiable content, such that the referent does not satisfy the content. Again the problem is Determination. If we assume Determination, and if the referent doesn't satisfy the content, the theory will not correctly predict it to it to be the referent.

Recall my case of Alice, who wrongly believed that Andrew Wiles had proved the Collatz Conjecture. According to a theory with Determination, Alice uses the name to 'refer' to the actual prover of the Conjecture (she only 'refers' because there is no such person—Alice actually fails to refer). But in fact she refers to the real Andrew Wiles, who proved Fermat's last theorem. When she uses the

name, Alice produces a token utterance of 'Andrew Wiles', and an A-judgement that the referent of the utterance proved the Collatz Conjecture.[75]

So the puzzle is 'solved' (really *dissolved*) in the same way in each case, just as expected.

### 4.3.2 Failure cases

These are also fairly straightforward. In a failure case there is a refeme with no referent, but which behaves in ways which are generally made intelligible by reference. Tokens of 'Feynman' are meaningful, and they are appropriate or inappropriate depending on their relation to their referent Feynman; 'Feynman was a physicist' is good (because he was), 'Feynman is alive' is bad (because he isn't). This all makes sense because there is a thing—the actual Feynman—to which the tokens relate. But tokens of 'Vulcan' are meaningful too: 'Vulcan does not exist' is true precisely because the token of 'Vulcan' does not refer. The same goes for cognitive cases: how can it be that I think about Vulcan when there is no Vulcan to think about? And in perceptual cases, there seem to be significant and contentful perceptual experiences, but without any object to account for this; hallucinating a pig is not the same as just not perceiving at all (or hallucinating a parrot), though in neither case does one *see* anything at all. The problem is to explain how all this works without appeal to a referent called 'Vulcan', or a real pig (or parrot).

Failure cases can be easily explained by my theory. The significance of a refeme is not exhausted by its referent, so it is entirely possible to produce a refeme with no referent, but which still has content; its content will be its ascription. On saying 'Vulcan does not exist', an agent produces an utterance of 'Vulcan' which does not refer, but they make an A-judgement which might be expressed as *the referent of the token I am now producing is a planet*, *the referent of the token I am now producing is between Mercury and the Sun*, etc. The speaker tokens this utterance, roughly because they believe that there is no planet between Mercury and the sun.

To be clear, this is not why the claim is *true*—it is true because it is a negative existential claim and the refeme does not refer (or equivalently, because Vulcan does not exist). If the speaker was confused, and thought that Vulcan was a real planet between Mercury and Earth (though had come by the name 'Vulcan' in the same way as usual), their utterance of 'Vulcan' would still not refer,

[75]If the conjecture is supposed to be a particular, such as an abstractum of some kind, then this is a case of impure ascription and the full story will include a further refeme for the conjecture; see §5.3.2.4 on apparent reference to abstracta. Alternatively, there may simply be a pure property *prover of Collatz Conjecture* which Alice ascribes.

and the negative existential would still be true, even though there *is* a planet which fits their ascriptions. Instead, the ascriptions of being a planet and being between Mercury and the sun make their utterance meaningful in a way which its referential status cannot, and distinguish such utterances from other non-referring ones. Thus an assertion that Vulcan does not exist is different from one that Pegasus doesn't.

A brief aside: this claim might sound to some like a version of descriptivism or another object-independent theory, at least with respect to the content of empty utterances in negative existentials and similar cognitive or perceptual phenomena. In fairly broad terms this is correct: the descriptivist insight is that the significance of non-referring tokens must be understood in terms of the properties their putative referents are supposed (by the tokens' producers) to have, not the actual referents or their properties (because there aren't any!). We might call this position *propertism*, because what is actually doing (some of) the work is a set of properties, rather than a structured linguistic entity like a description. My theory agrees with propertism. However, many descriptivists hold a much stronger claim: not only that the significance of an non-referring token must be understood in terms of properties, but that the token *itself* (or for Russell, the sentence) must be subject to analysis which will reveal a more structured entity, with elements which correspond to properties, and that *these properties* are the ones which explicate the significance of the original token.[76] Russell (1905) is a good example: the meaning (if any) of a token of a non-referring name like 'Apollo' is the contribution it makes to the underlying logical form of sentences in which it appears, and this contribution is a structure of quantification and predication specifiable only in formal, logical terms. Hence the name *descriptivism*: empty tokens apparently of a type which would usually gain their significance from their reference (i.e. empty 'names') are in fact not of that type at all (are not names). Instead they are, or can profitably be replaced by, descriptions, with elements which correspond to properties, which in turn explain the significance of the original token. This is a much stronger claim, and not one I am committed to. Propertism and descriptivism are often run together, but my theory supports only the former.

Cognitive failure cases work in the same way. To think about Vulcan is to produce a cognitive token which does not refer, with associated A-judgements that the referent of that refeme is a planet, and is between Mercury and the sun. These judgements make intelligible the agent's behaviour (their speech, their outward action, etc.).

[76]This explication is then generally understand as Determination, which I have rejected already.

Finally the perceptual failure cases. Hallucinating a pig, one tokens an object file which does not refer, but with associated A-judgements that the referent of that refeme is a pig. For the agent, it is as if they see a pig (which they would, if the refeme referred to a pig), but they do not. This is not a case of *misperception*, as where one tokens an object file which in fact refers to a dog, but one *seems* to be looking at a pig. Rather there is no referent at all.

As anticipated (and as with the satisfaction cases), the real challenge here is to offer a good theory of why an agent would token a non-referring refeme in the first place, and why they would make these associated judgements rather than others, but these are questions of determination, for the next chapter.

### 4.3.3 Frege's puzzle

The essence of Frege's puzzle is to explain why some identities are informative and others are not. A good way of making this plain is to contrast an uninformative case ('Hesperus is Hesperus') with an informative but otherwise similar one ('Hesperus is Phosphorous'); this is how Frege presents the case. In a presentation like this, there are four refemes overall: two in the trivial case, and two in the informative.

The puzzle can be presented differently. Millar's perceptual Frege puzzle compares two cases E* and E of different forms: in E* there is a visual presentation as of two objects, but in E there is a presentation as of only one, though in both cases there is only one object. In the case as Millar presents it there are only *three* refemes. I find the puzzle presented this way slightly harder to interpret, as the cases are less straightforwardly comparable. Nonetheless, the puzzle is the same, however it is parsed.

Since classic expositions of the puzzle in language use the four-refeme form, and since these are more straightforward to interpret, I will examine the problem in the four-refeme form first, and then describe my solution again, in the three-refeme form.

#### 4.3.3.1 Four-refeme form

Frege's puzzle in language has always been presented in terms of four refemes. In Trivial there are two tokens of 'Hesperus', and in Informative there is a third token of 'Hesperus' and one of 'Phosphorous'. Cognitive cases can work in the same way too: an agent who happens to be interested in the law of identity thinks to themselves that Hesperus is Hesperus, and in so doing produces two cognitive tokens. Then they think that Hesperus is Phosphorous (perhaps because they

have only just learned this, and must carefully remind themselves) with another two tokens.

Millar's perceptual version of the puzzle has only three tokens. My adaptation of Wettstein's (1986) case is a better comparison. In my case, you are backstage at a rock concert, hoping to catch the eccentric star. The star is made up and dressed such that their left and right profiles appear very different. They are also obsessed with self-image, and their dressing room is filled with large mirrors at odd angles. Standing in the room, you have an experience E which you take to be of two separate people, facing in opposite directions. In fact, you see the one star, strangely reflected in their own mirrors. Alternatively, if you were standing elsewhere, or were already aware of the star's bifurcated appearance, you might have an experience E*, presenting one side of the star, but reflected towards you in two separate mirrors, and which you thus took to be of one person reflected twice. Experience E* presents nothing which E does not, and the two situations involve exactly the same actual objects. Yet E and E* differ significantly in their content.

In my perceptual case, there are four refemes: one in each mirror in each of E and E*, and all four refer to the same person.

Now for the solution. I said above that the theoretical role of ascription is to describe what agents actually *do*. Generally, when things go as they should, we ascribe to an object only properties it actually has (and ideally, we do so *because* it has them—more on this in the next chapter).

Millar describes Frege's puzzle as a mistake of one object for two and this way of talking may be illuminating here. In the informative cases there is really only one object, but the agent behaves as if (i.e. they *ascribe* as if) there are two: Frege's puzzle involves co-referring refemes which are not treated as if they are co-referring. So: the agent produces two refemes, which in fact co-refer, and to which therefore the agent *should* ascribe the same things, but they do not.[77] In the linguistic case (i.e. Trivial), these refemes are utterances of a name. In the cognitive case they are cognitive tokens, and in the perceptual case they are object files. In my case of the vain rock star, the agent tokens one object file for one reflection, and another for the other, but ascribes very different things to the referents of the two files in spite of their actually co-referring (surely because they look so different).

In the trivial cases, the agent makes no mistake, because they correctly treat the two refemes as co-referring: they ascribe the same things to both. It should

[77]The principle I am relying on here is that when two refemes co-refer, the agent should ascribe the same things to each. The converse does not hold though: an agent might perfectly correctly ascribe the same things to the referents of two different refemes, for example if there are two different objects about which they know exactly the same things.

be clear by now how this works in the linguistic and cognitive cases. In the rock star scenario, the agent again tokens one object file for one reflection, and another for the other, but ascribes the same things to them both (and surely for the same reason—that they look very similar).

This explains the important difference between the cases: in one, the contents match, and in the other they do not. Frege's cases are special in that both sets of different ascriptions in the latter case are satisfied by the object in question: Venus (Hesperus, Phosphorus) both rises in the morning and in the evening; the rock star really does look *this* way from the right and *that way* from the left. So when we learn that the identity claim between the two refemes is true, the case can 'contain very valuable extensions of our knowledge' (though exactly how this extension works is not the concern here).

#### 4.3.3.2 Three-refeme form

This solution can be straightforwardly transposed into the three-refeme form. In Millar's perceptual case, there are three refemes: one involved in an experience E as of a single, whole dog, and two involved in a second experience E* as of two dogs, each partially obscured.

In the four refeme form, the 'mistake of one object for two' is in ascribing different things to the referents of two actually co-referring refemes, and the contrast is with a non-mistaken case where the same ascriptions are made to two further co-referring refemes. In the three-refeme form, the mistake is the same, but the contrast is with a case where the mistake cannot arise at all, because only one refeme is tokened (it is impossible to fail to identify an object with itself).

In E* the agent tokens two refemes, which in fact co-refer, though the agent ascribes different things to the putative referents of each. They might ascribe *having a visible snout* and *having an occluded tail* to one, and *having a visible tail* and *having an occluded snout* to the other. Since the refemes actually co-refer, the agent 'should' make the same ascriptions, but they don't. In E the agent tokens only *one* refeme, and makes ascriptions to its putative referent. In this case it is not even possible that the ascriptions not match, because there is only one refeme at hand.

#### 4.3.3.3 'Paderewski' cases

Recall the second form of the 'Paderewski' cases introduced in §2.2.1 and often discussed alongside Frege's puzzle. Peter is aware of both Paderewski the pianist and Paderewski the statesman, though believes wrongly that they are distinct, and reminds himself of this by saying:

**Peter's Denial** *I must remember: Paderewski is not Paderewski.*

This is just a case of Frege's puzzle. As with Informative, there is really only one object, but the agent behaves as if—i.e. *ascribes* as if there are two. There are also some minor differences.

Peter's Denial differs from Informative in two ways. First, it uses identical terms on both sides of the claim, where Informative uses different terms. Second, it is a denial of identity, where Informative is an assertion.

We have seen already the that bearers of content, and therefore of reference, are not terms but individual utterances *of* terms. So there is a sense in which the first difference is merely apparent: in Informative, and in Peter's Denial, there are two different linguistic refemes—the first utterance of a term, and the second. It so happens that in Peter's Denial they are tokens of the same term. But since there are two distinct refemes, it is entirely possible that they have distinct contents; this is precisely the similarity between Peter's Denial and Informative. Since the claim is non-trivial, and since the two refemes co-refer, they must differ in their ascriptions. And this is exactly what we should expect: uttering the first token of 'Paderewski', Peter ascribes *being a pianist* (but not *being a statesman*), and uttering the second ascribes *being a statesman* (but not *being a pianist*). These ascriptions are all correct, but because Peter doesn't realise that the pianist is the statesman, he does not make both ascriptions for both utterances.

This demonstrates the importance of the second difference between Informative and Peter's Denial—it is important that Peter's Denial be a *denial* of identity. It is occasionally suggested that upon realising the identity of the pianist with the politician, Peter might suddenly say 'Paderewski is Paderewski!' and express a non-trivial true identity claim, as a positive equivalent of Peter's Denial.

The analysis seems straightforward. Peter has suddenly realised that the pianist is the politician. Wishing to express this, he utters 'Paderewski', referring to the man and ascribing only *being a pianist*, then 'is', and finally 'Paderewski' referring to the same man, but now ascribing only *being a statesman*. There would be two co-referring refemes (true identity claim), but with different ascriptions (non-trivial).

But no. If Peter really has realised that the politician is the pianist, then surely upon intentionally referring to the one man (whom he now knows to be one man with both properties), he will ascribe both *being a statesman* and *being a pianist*. He will do this in both utterances of the name 'Paderewski', and thus their ascriptions will not differ in the expected way. This line of reasoning can be followed for any difference in ascription posited to explain the difference

between the first and second tokens of 'Paderewski' in the positive version, so attempted positive claims can never be of the same kind as Informative.

Peter's Denial doesn't have this problem because Peter does not realise that the pianist and the statesman are one, and therefore will not ascribe both properties when uttering either token of the name. If he did, the case would be impossible. So the second difference between Peter's Denial and Informative is not merely an accidental feature of the example, but in the nature of the case.

This then is the analysis of the second form of the 'Paderewski' case, in so far as it is relevant here: a single subject may make two co-referential utterances of the same name, though they make different ascriptions in the two utterances, because they do not realise that the utterances co-refer. Denials of identity of the form ⌜*a* is not to *a*⌝ provide a stark example. 'Paderewski' cases are just an instance of Frege's puzzle.

### 4.3.4 Transformation

In Chapter 3 I described another puzzle: a magician waves a king in the air and before your eyes it 'becomes' an ace. Similar tricks are done with different coins, a cane and a scarf, etc. The explanation here is clear: there are two objects with only one refeme between them, whenever one is visible the other is not, and the case is set up so that the single refeme refers to whichever *is* visible. Because refemes are only 'supposed' to track single objects over time, it appears that a single object is present throughout. But the object at the beginning and the object at the end have very different properties (one looks like a king, the other an ace) so the refeme's ascriptive content changes as one is replaced with the other. From the agent's point of view it appears that one object which began looking like a king now looks like an ace.

## 4.4 Conclusion

Let us take stock. This chapter described my theory of the nature of content. The theory was simple: a refeme's content is the conjunction of its ascription and its referent (if any). We have seen that these are much less closely coupled than might be thought, and how this theory provides 'semantic' solutions to the motivating puzzles. These solutions had a constant form across linguistic, cognitive and perceptual versions of the puzzles, which counts as further evidence for the original Unity claim.

I have also left two promissory notes for the next chapter: first that I will explain the general mechanism by which each refeme has the content it does

(the 'metasemantics') and second that this will explain why the various puzzling cases come about in the first place.

# Chapter 5

# The Determination of Content

Which refemes have what content? This is the question of the *determination* of content, and it is the topic of this chapter.

In Chapter 4, I provided solutions to the puzzles which originally motivated this dissertation (the same puzzles upon which I grounded my argument for Unity in Chapter 2). But as I said there, the solutions I gave were rather thin. The framework I have developed so far will tell us what the bearers of content are in each case, and what their content is. But in many of the puzzling cases it is not clear *why* a situation with such refemes having such content would arise in the first place. A theory of the determination of content will tell us.

I have argued already that content includes both reference and ascription, and that the latter does not determine the former. So the questions of determination for reference and ascription come apart. Really then, I must offer two separate theories: one of reference-determination, the other of ascription-determination. Although I do offer two theories, we will see below that the two issues are intimately related.

My theory of reference-determination is inspired by 'causal' theories of name-reference given by Kripke, Devitt, Chastain and Sainsbury.[78] They say that the reference of an utterance of a name is fixed by a causal chain leading back to some initial reference to an object. So there are two parts to the theory: first, the initial or 'primitive' reference (in the theories of names this is often associated with a 'baptism' or 'dubbing' event), and second, the inheriting, 'borrowing' or

[78]The origin of such theories is generally attributed to Kripke (1980). The history is definitely more complicated. Fetzer and Humphreys (1998) offer a survey of then-recent debates, especially concerning Ruth Barcan Marcus' contributions. Several authors mention earlier work by Peter Geach (1969; 1972), and Gentry (1943) attributes a causal theory of reference to Ogden and Richards (1923). Their work includes a clear articulation of the idea of a 'chain' of reference (p. 11) using the example of 'Napoleon'. The fact that the same name was used by Kripke in his famous sketch of a causal theory may indicate that he was inspired by their work, or it may be merely a coincidence. Burch (1979) argues that William James had originated a similar theory by 1885.

'deferring' of reference by subsequent utterances. The causal theories add that the relation by which later utterances inherit from earlier ones is causation.

I deny this last claim, and several others common to the causal theorists, but my theory of reference-determination retains the primitive/inherited structure, and generalises it to non-linguistic refemes. That is: some refemes get their reference 'primitively' by a relation with the referent, whereas some 'inherit' it by having an appropriate relation with another refeme, which itself refers to the referent. My detailed account of the nature of initial reference is one of the principal improvements of my theory over earlier ones.

I extend the primitive/inheritance model to ascription as well: some refemes get their ascriptive content by having a certain relation with another refeme, which itself had that ascriptive content. Of course, it cannot be that *all* refemes get their ascriptive content determined this way, on pain of infinite regress. Just as there is primitive reference-determination at the end of the reference-inheriting chain, there must be 'primitive' ascription-determination at the end of the ascription-inheriting chain.

Just as linguistic, perceptual and cognitive refemes have the same kind of content, I argue in §5.1 that *what* content each has is determined in the same way; not only do I claim that the nature of content is unified, but that so is the mechanism by which it is determined. It is an advantage of my theory that there are links of inheritance, for reference and ascription, between refemes *of different kinds*. This provides a clear and systematic explanation of how particular instances of language, thought and perception interrelate. I believe mine is the first account to develop this idea in detail.

I initially introduced the primitive/inheritance model as a way of understanding reference. In the theory below however, my account of reference-determination depends on details of that of ascription-determination. Accordingly, we start with ascription in §5.2: §5.2.1 is on the primitive/inherited model, §5.2.2 on primitive ascription and §5.2.3 on the mechanism of inheritance. §5.3 is on reference and follows a similar structure. With all this in hand, §5.4 delivers the promised 'metasemantic' solutions to the original puzzles.

## 5.1 Whence a unified theory of determination?

I argued in Chapter 2 that we need a unified theory of the nature of content. And after some ground clearing in Chapter 3, I provided such a theory in Chapter 4. But the resources this theory provided were rather thin in comparison with the problems it was supposed to solve, and in the present chapter a theory of determination is given to make up the ground.

All very well, an objector might say, but why only one theory? It was a substantive claim that the nature of content is unified and we could provide a single theory; it is a substantive claim that determination is unified too. If it isn't, we will need at least three theories to cover language, perception and thought. Utterances of names, cognitive tokens and object files are very diverse kinds of thing, why think that the mechanism by which one token gets its content is the same as that by which another gets *its* (even if the content so gotten is of the same kind)?

My answer comes in two parts. First, a straightforward reply as to why we should expect the theory to be unified, though the kind of unity it justifies may be too weak for some readers to think the term properly applied. Second, a caveat to this, arguing that this limited kind of unity is exactly what we should expect.

First, the straight answer. In many cases, it is clear that a certain token has the content it does because of its relation with some other token, and that latter token's properties. I do not claim that this is always the case; sometimes it works differently. But this happens often.

For example, it is widely agreed that for many historical figures' names, our contemporary ability to use the name to refer to the person is linked to some event of someone's perceiving that person. In a counterfactual scenario where no such perception occurred, but everything else went as it really has, our modern utterances would not refer to the person. Similarly, many authors have thought that hearing or understanding an utterance of a name endows the hearer with the ability to think about the utterance's referent, even if they know very little about it.

My point here does not require any committed view on how these links work. All I am pointing out is that often, what content some token has is intimately related with what content some *other* token of a different kind has or had.

So providing a theory of content-determination for only one kind of token (or even separate theories for each) would be painting half a picture. One may as well answer 'how are you able to refer to Socrates?' with 'because I have read Plato'; not a false answer, but not a full one either. Instead, a good theory of content-determination will include these inter-kind links *in the theory*—a unified theory of content.

As advertised, this is a rather thin kind of unity. The claim is not really that the mechanics of determination are exactly the same for all different kinds of refemes. Rather it is that for all kinds of refemes so far considered, any full account of the mechanics of their determination must include some account of

the mechanics of determination of *other* kinds of refemes. So a theory for any is a full theory of all.

Now for the caveat; this is just the kind of unity we should expect. Obviously, a complete and true theory of what determines the content of a perceptual token will be different from one for a cognitive or linguistic token. At the very least the two must differ in certain constraints: we can only perceive perceptible things, but are perfectly capable of thinking or talking about imperceptibles like dead relatives or occluded objects. If there is any sense in which the mechanism of determination is unified across the different domains, then it is that there is a single holistic theory encompassing their diversity, not that there is one theory, abstract from each, which is true of them all.

With this out of the way, let us turn to the theory proper.

## 5.2 Ascription

### 5.2.1 The primitive/inheritance model

Some refemes get their ascriptive content directly or 'primitively'. Some get it by 'inheritance', by being appropriately related to another refeme with that same content. This section argues for these two claims.

One often produces a token with certain ascriptive content just because one has previously produced a different token with that same ascriptive content.

Earlier today I met someone for the first time, with a noticeably big nose. I produced a cognitive refeme referring to them, and ascribed *having a big nose*. Later I see someone from behind, walking along in front of me. I token a cognitive refeme R referring to them. From this angle their face is obscured by their large coat and voluminous hair. Nonetheless, because I believe they are the same person as I met earlier, I ascribe again *having a big nose*. I thus make an A-judgement concerning the referent of R that it—whatever it is—has a big nose.

The 'because' of the story is telling. It describes why my eventual A-judgements include *having a big nose* and not some other characteristic: because my earlier A-judgements did. If my earlier ascription had been of *having blue eyes* (and not a big nose) then my present ascription would be too. This case is typical—many refemes have the ascriptions they do because of a relation with another refeme which also had that ascription. Some refemes inherit their ascriptions.

But it cannot be turtles all the way down. Even if some refemes do get their ascription by inheritance, eventually the chain must end with some refemes which get their ascriptions 'primitively'. We have arrived at the primitive/inheritance model.

It is important to note that 'primitive' and 'inherited' are not two different kinds of ascription. All ascriptive content is of the same kind. 'Primitive' and 'inherited' are two different ways by which a certain refeme might come to have the ascriptive content it does. Even if Alice earned her millions as a trader and Bob inherited his from his grandmother, their money is worth just the same.

### 5.2.2 Primitive ascription

I have relatively little to say about the exact mechanics of primitive ascription.

A refeme gets its ascriptive content primitively when, as a result of other properties or functional roles of the refeme itself, the agent makes A-judgements about the refeme's putative referent. The exact (probably subpersonal) mechanics of how this happens is a question for the cognitive scientists, and perhaps those philosophers interested in the nature of qualia, but not for me.

For example, suppose I seem to see a black coffee cup. I token a visual refeme which tracks the apparent cup in my visual field. My visual experience also includes various colour qualia (various shades of blackness and greyness, etc.). As a result of these qualia, I form A-judgements about the putative referent of the refeme; I A-judge that it—whatever it is—is black.

The important point is that these A-judgements are not the result of a relation with *other* refemes and their ascriptions. Such A-judgements would be *inherited* ascriptions, and are described in the next section.

It is a consequence of this theory that only perceptual and cognitive refemes have primitive ascription.

### 5.2.3 Inherited ascription

In the example above, I described how one refeme had certain ascriptive content (*having a big nose*) because of its relation with another refeme which also had this content. What is this relation? In the example above, the contemporary refeme inherited its content from the earlier refeme because the agent thought that they had observed the same object twice, and took evidence from the first sighting to apply in the second. The process could be caricatured: I saw someone earlier and they had a big nose; I am seeing the same person now; the person I see now has a big nose (importantly this is really *just* a caricature—I am not committed to any such explicit or conscious process of reasoning).

This case is not only an example, but a paradigm: one refeme inherits its ascriptions from another just when, and just *because*, the agent producing the former refeme believes they co-refer (modulo some wrinkles below).

Here is a more detailed model. I have already claimed that when we produce a refeme, we make judgements. Some of these judgements, which fulfil further criteria, are *ascriptions*. Now I claim that there is another special kind of judgement besides ascription. When an agent produces a refeme, they have zero or more associated judgements which can be expressed in natural language as an instance of the following schema:

**Co-Reference** *The token I am now producing co-refers with a certain previous token.*

That is, when I produce a refeme R1, sometimes I judge that it co-refers with some other refeme R2 (which I or another agent might have produced). Let us call these 'CR-judgements'. As with Ascription, the natural language form of Co-Reference is merely illustrative; I do not claim that CR-judgements must themselves have sentential or linguistic form. As with A-judgements, CR-judgements may be more or less tacit or explicit, though most are generally more tacit. And as with A-judgements, my theory does not place any restrictions on which CR-judgements can be made (even if some would be irrational, improbable or otherwise a bad idea); in particular, CR-judgements may be mistaken. Finally, as with A-judgements, the apparent references to 'the token I am now producing' and to 'a certain previous token' are to be interpreted as a constitutive relation between the CR-judgement and a particular refeme or refemes, which I leave primitive and unanalysed.

There are two cases of CR-judgement to consider. First where the present refeme has primitive ascription. In this case, the CR-judgement is just the same as it seeming to the agent that they recognise an object (the putative referent of the present refeme) as one referred to previously (the putative referent of the other refeme). For example, in the morning I see a robin sitting on my windowsill; I token a visual refeme which refers to the robin and has certain ascriptive content. Later, I token another visual refeme and (for whatever reason) make a CR-judgement about the later refeme and the earlier refeme; this is just a jargonistic way of saying that I judge them to co-refer. My claim about recognition is that, if the present refeme has primitive ascriptive content, then the jargonistic description is *also* another way of saying that I seem to recognise the robin I saw earlier. Of course, whether I actually *do* recognise the robin depends on whether it actually *is* the one I saw earlier, but this is not the issue here.

I characterised this first kind of case as where the *present* refeme has primitive ascription. In the case of the robin, both refemes have primitive ascription, since both are perceptual, but this need not be so. My partner tells me they saw a white raven, and I produce a cognitive refeme which I take to refer to it. Later

that day I see a white raven. I token a visual refeme with ascriptive contents *being a raven* and *being white*. Finally I make a CR-judgement about the later refeme and the earlier refeme; this is just a jargonistic way of saying that I judge them to co-refer, or that I take myself to be seeing the thing I thought of earlier, after my partner told me about it. Here, I seem to recognise an object as one referred to earlier, though the time of recognition is the first time *I* perceive it. Again it is irrelevant whether I really *am* seeing the thing they told me about (indeed it is irrelevant whether there is any such thing to be seen).

In the second case of CR-judgement, the present refeme does not have any primitive ascriptive content. In this case, I claim, the present refeme takes on whatever ascriptive content which the agent presently judges the earlier refeme to have had and which is not ruled out by other ascriptive content.

Let's focus on the first condition. A refeme R1 without primitive ascriptive content, and CR-judged to co-refer with some other refeme R2, has whatever ascriptive content R2 is presently judged to have had. In more natural language (and reflecting several of the examples above): if one takes oneself to be referring to the same object twice, one will take it that the object has the same properties now as (one now judges it was judged to have) then.

This condition captures the kind of inheritance illustrated in the examples above. In a case of inherited ascription, the agent judges that two refemes co-refer. From the agent's point of view then, other things equal, anything true of the referent of one refeme should be true of that of the other. The condition just says that whatever the agent presently believes was judged true of the putative referent of the other, they in fact judge true of the putative referent of the one now. Moreover, they do so precisely *because* they believe that the two refemes co-refer.

This mechanism explains how cognitive tokens and name-utterances have ascriptive content based on the prior perceptual experiences (or cognition or speech) of an agent. I see a sparrow and token a visual refeme R1 with primitive ascriptive content including *being a sparrow* and *being brown*. Then I think about the sparrow I seem to see: I token a cognitive refeme R2. R2 has no primitive ascriptive content, instead it gets its ascriptive content by inheritance: as I token R2, I CR-judge that R2 co-refers with R1, and I judge that R1 has/had ascriptive content including *being a sparrow* and *being brown*. Thus, I A-judge of R2 that its referent—whatever it is—is a sparrow, and similarly that it is brown. If I token a name-utterance R3 and CR-judge it to co-refer with R2 (and correctly judge that R2 ascribed *being a sparrow* and *being brown*), then R3 will also have the ascriptive content *being a sparrow* and *being brown*, and for similar reasons.

A similar process occurs with ascription-inheritance between name-utterances, though unlike the vision/cognition case, there is often a substantial lag between the two utterances. At T1 I produce an utterance U1 of the name 'Napoleon' and A-judge of the referent of U1 that it is an emperor.[79] At T2 I produce U2, another utterance of 'Napoleon', and CR-judge that U2 co-refers with U1 (let us suppose that this is true, though my account of reference only comes in below). At T2 I judge that U1 ascribed being an emperor, and so as I produce U2, I A-judge that the referent of U2—whatever it is—is an emperor.

When things go as they should, which they generally do, an agent will make correct CR-judgements about different refemes and correct judgements about the ascriptive content of the former refeme as they produce the latter. In the examples so far, everything goes as it should. But this isn't always so, and some detailed explanations of failure may help clarify or motivate my account.

First, agents make false CR-judgements. It often happens that one tokens a refeme, thinking it co-refers with some earlier refeme, though in fact it does not. Suppose I join an ongoing conversation involving many uses of the name 'Napoleon', and hear that 'Napoleon studied at the École polytechnique in Paris'.[80] I correctly judge that this utterance of 'Napoleon' ascribes the property *studied at the École polytechnique*. Unknown to me, the man under discussion is Napoleon Daru, an obscure French soldier and politician. Wrongly assuming that the discussion concerns Napoleon Bonaparte, I token my own utterance of 'Napoleon' referring to Bonaparte, saying 'ah yes, and later Napoleon fought in the war'. In fact Napoleon Bonaparte studied at the École Militaire, though I am unaware of this. Since I judge (correctly) that the earlier refeme ascribed *studied at the École polytechnique* and CR-judge (mistakenly) that the earlier refeme co-refers with my present one, with my present refeme I ascribe *studied at the École polytechnique*. The inheriting of ascription is perfectly correct in that I have correctly identified some of the ascriptive content of the early refeme, and copied it onto a refeme which I CR-judge to co-refer. The problem is that the copying is inappropriate, because the two refemes do *not* co-refer.

Secondly, agents sometimes make wrong judgements about the ascriptive content of the former refeme as they produce the latter and so the refemes they

[79]There are many reasons this might happen. In particular, it might be because I CR-judge my utterance to co-refer with some yet earlier utterance (and so on ad infinitum). But we need not fill out those details in this example.

[80]To be clear: the fact that the utterance of 'Napoleon' ascribes this property has nothing to do with the fact that the same speaker who utters 'Napoleon' then immediately utters 'studied at the École polytechnique in Paris'. This latter utterance does not ascribe anything to Napoleon, because ascription is a matter of the mental state of the speaker at the time they utter 'Napoleon', not of later utterances. However, it is by hearing this later utterance that I, the hearer, am able to correctly infer what was ascribed in the utterance of 'Napoleon' just before it, namely the property of having studied at the at the École polytechnique in Paris

produce can 'mis-inherit'. Consider another variant case: at T1 I produce an utterance U1 of the name 'Montgomery' and A-judge of the referent of U1 that it is a famous general. A long while later at T2, I produce U2, another utterance of 'Montgomery', and CR-judge that U2 co-refers with U1 (again, suppose that this is true). But at T2 I wrongly judge that U1 ascribed *being a famous spy*, and so as I produce U2, I A-judge that the referent of U2—whatever it is—is a spy. Here I have correctly identified a pair of refemes between which inheritance 'should' happen. The problem is that I am mistaken about *what* should be inherited.

In both problematic cases, I end up ascribing the 'wrong' properties, in that the referent and ascriptions of my refemes are mismatched. But this happens for different reasons in the two cases.

Now the second condition. I said that in cases of CR-judgement where the present refeme does not have any primitive ascriptive content, it takes on whatever ascriptive content the agent presently judges the earlier refeme to have had and which is not ruled out by other ascriptive content. What is this 'ruling out'?

Sometimes an agent withholds an ascription they would otherwise make (given the mechanism so far described) as inappropriate. Suppose I see a young child one day and token a visual refeme with ascriptive content *being young*. If years later I token a cognitive refeme which I take to co-refer with the visual one, and if I am sufficiently aware of human biology and the passage of time, I will not ascribe *being young* to the putative referent of the contemporary refeme, even though I judge correctly that the earlier refeme ascribed this, that that ascription was correct, and that it co-refers with the contemporary refeme. I withhold the ascription because I recognise that in this case, 'copying' the ascription from one refeme to another is a bad idea. In other words, the ascription is 'ruled out' for the agent.

This might happen for various reasons. In the child case, the agent believes the ascribed property to be of such a kind that if the ascription was appropriate before, it cannot be now (as it happens, the agent is right, but the case doesn't require this). In other cases, the agent may believe the earlier ascription was never true to begin with. My friend tells me of the clairvoyant they met, 'Madame Arcati'. Believing they really did meet a Madame Arcati, but not that she speaks with the dead, I token a cognitive refeme which I CR-judge to co-refer with my friend's utterance of 'Madame Arcati', though I do not ascribe *able to speak with the dead*, even though I judge that my friend did ascribe this when they tokened their utterance. In this case the agent is right—nobody speaks with the dead—but this isn't necessary for the case. The converse might happen: my friend tells me they met 'Ajay Kakkar', a member of the Order of the Garter.

Believing they really did meet an Ajay Kakkar, but not that he is a member of the Order of the Garter (because we just don't meet people like that) I token a cognitive refeme which I CR-judge to co-refer with my friend's utterance of 'Ajay Kakkar', though I do not ascribe *member of the Order of the Garter*, even though I judge that my friend did ascribe this when they tokened their utterance. In fact I'm wrong, and my friend really did meet Ajay Kakkar, eminent surgeon and member of the Order of the Garter. There may be more reasons besides.

The second condition on ascription inheritance just says that properties ruled out like this, by the agent 'withholding' ascription for any reason, are not ascribed—even when they *would* be ascribed according to the rest of the account presented so far. Importantly, the 'ruling out' is a fact about the agent and their beliefs—not whether the relevant property does or doesn't obtain of the relevant object. In the Ajay Kakkar case, what rules out ascribing *member of the Order of the Garter* is not whether Kakkar actually is or isn't a member, but whether I believe that the referent—whatever it is—of my friend's utterance is a member.

It is for this reason that my formulation above just excludes any ascriptions which are 'ruled out' in this way. The ruling out might be for various different reasons, and might or might not be 'good' depending on the agent's faculties. None of this is within the purview of my account. All I claim is that properties screened off in this way are not ascribed.

#### 5.2.3.1 Multiple inheritance, conflict and dominance

There is a shortcoming in the account so far. Consider the sparrow case again. In the version above I token a visual refeme R1, then I token a cognitive refeme R2 which I CR-judge to co-refer with the visual refeme and finally I token a linguistic refeme R3 which I CR-judge to co-refer with the cognitive refeme R2. Ascription gets passed along this chain of co-reference. Surely in a real version of such a case I would also CR-judge R3 to co-refer with R1. So R3 has two sources of inheritance: R2 and R1.

The first thing to make clear is that my account is compatible with such cases. There is nothing to say that they can't or shouldn't happen, and the description in terms of multiple inheritance is the most natural way of understanding them given the terminology I have developed so far. Indeed, the *vast* majority of actual cases of ascription inheritance will include multiple inheritance.

There is a problem though. The problem is not with the diffuse structure of inheritance in these cases, but that there may be gaps or inconsistencies between ascriptions inherited from different sources.

A 'gap' is where one refeme inherits ascribing F from another (which is silent on G), and ascribing G from yet another (which is silent on F), and the agent does not rule out F and G co-occurring. One colleague tells me they have a student 'Jack', who moonlights as a circus performer. Another colleague tells me *they* also have a student 'Jack', who is a good student and participates actively in seminars. As a consequence of these two episodes, I token a single cognitive refeme, which I CR-judge to co-refer with both my first and my second colleagues' name-utterances. Correctly judging that the first colleague's refeme ascribed *being a circus performer* would lead me to ascribe that property, and similarly for my second colleague and *being a good student*. And indeed, this is exactly what happens: I end up ascribing both properties to the putative single referent of my cognitive refeme. The ascriptions I inherit from one colleague's refeme cover the 'gap' left by the silence of the other's.

An inconsistency is where one refeme inherits ascribing F from another (which is silent on G), and ascribing G from yet another (which is silent on F), and the agent *does* rule out F and G co-occurring.[81] I see a bird in the morning, token a visual refeme V1 and ascribe *being a swallow*. I see a bird in the afternoon, token a visual refeme V2 and ascribe *being a swift*. Later that evening, I reflect on the day's events and (for whatever reason) I token a single cognitive refeme C3 and CR-judge both that C3 co-refers with V1 and with V2. As a rational agent and a good birdwatcher, I withhold the possibility of ascribing both *being a swallow* and *being a swift*. Cases like this are problematic because the agent:

1. CR-judges that their contemporary refeme (C3) co-refers with some earlier refeme (V1).
2. CR-judges that their contemporary refeme (C3) co-refers with some other earlier refeme (V2).
3. Judges that the putative referent of the former earlier refeme (V1) had the property A (*being a swallow*) (because they judge, now, that the refeme V1 correctly ascribed A).
4. Judges that the putative referent of the latter earlier refeme (V2) had the property B (*being a swift*) (for similar reasons).
5. Withholds present co-ascription of A and B (nothing is both a swift and a swallow).

[81] A special case of inconsistency is where $G \equiv \neg F$, and the agent rules out co-occurrence because they have a basic rational belief that nothing can both instantiate a property and its converse. I leave this case implicit because rationality is only one reason that a certain combination of properties might be ruled out, and because I am suspicious of the philosophical currency of negative properties like ¬F, even to the extent that I am suspicious of *thoughts about* ¬F (for an overview on negative properties see Orilia and Paoletti 2025 §4.1).

Just as with mis-inheritance earlier, it is irrelevant here whether the earlier refemes actually *did* ascribe such content, or whether either of the earlier refemes in fact referred to an object which in fact had the properties A or B.

Withholding any one of the five judgements makes the resulting set consistent. So there are five ways the agent can 'resolve' the case, corresponding with the five different judgement-sets produced by removing one judgement from the set above. The elements of any such set will be mutually consistent.

The major claim of this section is that when actually faced with such a case, the agent will select one of these sets by a deciding principle of 'preference' or 'dominance', which is an idiosyncratic element of their psychology (and perhaps circumstances). That is, the agent is such that they prefer to ascribe certain properties in certain situations, or that *for them* certain ascriptions 'win' over others in conflicts. As with making A-judgements and CR-judgements, this process of 'deciding' need not be explicit, or even consciously accessible to the agent. And as with the explanations of why we make certain CR-judgements and not others, explanations why certain agents have certain preferences or dominances are not at issue here. There are many reasons, I will only suggest an illustrative example.

In the swallow/swift case above, I might just think it more likely that I saw one swallow twice (versus having seen one swift twice or two different birds), and thus withhold my judgement that the putative referent of the latter earlier refeme (V2) was a swift. I thus end up with mental states committed to my having seen a single swallow twice.

Importantly, the judgement I withhold is my present judgement about the putative referent of the refeme V2: I hold back from judging *now* that it—whatever it was—was a swift. I might also make a present judgement that the refeme V2 ascribed *being a swift*; I need not withhold this. Even if I now judge that the referent of V2 was not a swift, whether it was or not does not affect whether I judged *at the time* that it was a swift (and thus should not affect how I judge now on whether I so-judged then).

## 5.3 Reference

In the three sections below, I give an account of what determines which object, if any, a refeme refers to. This is very narrow project. It is not a theory of perception, or of meaning or truth, or any other such thing, though I will draw on literature on these topics. It is not even a theory of reference, in one sense of that phrase. I do not develop a theory of what reference *is*, but of how it is *distributed*—the general rules about which things refer to which others. Of

course, one might believe that this is all there is to a theory of what reference is (a sort of functionalism about reference) but I am not concerned with that claim.

My ideas are influenced by 'causal' theories of reference, in three important ways. First, causal theories are an instance of the primitive/inheritance model. I argue in §5.3.1 that reference instantiates this model. Second, it is widely thought that a causal relationship between referent and refeme is a necessary condition on primitive reference. (e.g. Grice 1961) Though this is attractive, it is surprising how little developed theory there is of exactly how primitive reference works, or how it 'kicks off' the ongoing chain. I develop a fuller account of primitive reference in §5.3.2. Third, causation is often taken as a reasonable candidate for the relation between two refemes such that one inherits its reference from the other. I give my account of reference inheritance in §5.3.3 (though causation has an important place here, I do not hold that causation is always necessary for reference inheritance).

The second and third claims suffer from the problem of 'proliferation': there are just too many things in the causal history of any refeme for causation to fix reference, either primitive or inherited. In the primitive case, a huge variety of different objects and events are part of the causal history of any refeme, but at most one counts as its referent. In the inherited case, most refemes have a great number of other refemes in their causal history, between them referring to many different things, but the refeme usually only ends up inheriting one referent[82] (Cf. §5.3.3.4 on cases where the refeme does not only inherit one referent).

My accounts of primitive and inherited reference mostly consist in solutions to their respective problems of proliferation. Each section begin with a discussion of the role of causation: I argue a causal link is necessary for primitive reference, and though not necessary, is of one of several 'I-relations', where an I-relation is necessary for inherited reference. I then present solutions to the proliferation problems. Both solutions involve specifying further conditions on top of I-relation, which are also necessary for reference, and which with I-relation are jointly sufficient for it.

I have claimed that reference is only one aspect of the content of refemes. The other aspect is ascription. My solutions to both problems of proliferation involve conditions on a refeme's ascriptive content.

This might seem surprising. It is widely agreed that the proliferation problem for inheritance is not solved by adding a condition that the object must satisfy the relevant ascriptions. This is exactly what non-satisfaction cases demonstrate.

[82]The most famous example of this problem is Evans' (1973) case of 'Madagascar', but many others have noted something similar: one reading of Kripke (1980 p. 96) on Napoleon the Aardvaark, some comments by Gentry (1943), Stampe (1977 Sections 4 and 5.11) and Fodor (1984 p. 240) etc.

I have cited many such cases already and gave my own examples in earlier chapters.

But cases like this only apply to the problem of proliferation for inherited reference; the problem for primitive reference remains an open issue. So it is prima facie plausible that primitive reference *does* require satisfaction of ascriptive content. I defend just this view in §5.3.2. I claim that as well as a causal relation, primitive reference fixation requires that a refeme's ascriptive content be significantly satisfied by an object, and that this satisfaction is robust under certain kinds of change.

A great advantage of the primitive/inheritance structure is that the conditions for primitive and inherited reference come apart. On this model, even if primitive reference determination requires the satisfaction of ascriptions, inherited reference determination need not. The cases of alternative- and non-satisfaction demonstrate convincingly that it does not, and so my theory of reference inheritance does not require ascription satisfaction, though ascription does play an important role.

First I argue that causation is not necessary for inheritance. I present cases which clearly exhibit the primitive/inherited structure I have motivated, but where the relation by which the inheriting tokens inherit their reference cannot be causal. This is not always the case—some tokens clearly do inherit their reference along causal lines—but it is sometimes. Accordingly, I will offer a more general theory of the kinds of relation which can support reference inheritance (§5.3.3.1). I argue that different kinds of refemes are part of different 'refeme systems' and that each system comes with an 'I-relation' along which reference is inherited. So where a classical causal theory claims two refemes have to be causally related for reference inheritance, I instead claim they must be I-related. Causation is one such I-relation, so I retain much of the original intuition.

Secondly, I argue that for one refeme to inherit its reference from another, the agent producing the inheriting token must take them to be members of a common kind (§5.3.3.2). In the linguistic case, they must be taken to be utterances of the same name, in the cognitive case tokens of the same concept, etc..

Finally, I argue that for one refeme to inherit its reference from another, the ascriptive contents of both refemes must significantly *overlap*, regardless of whether either is satisfied by their actual referent (§5.3.3.3).

Each of these conditions is necessary for reference inheritance. Together they are jointly sufficient.

### 5.3.1 The primitive/inheritance model

The primitive/inheritance model applies to the determination of reference too. I will argue for this from two angles. First, a negative argument, that among the various existing theories of reference, a causal one is the best (causal theories have just this primitive/inheritance structure). Then I offer a positive argument, that given the work of the earlier chapters, causal theories explain reference-determination in exactly the way we would expect.

The negative argument. It is generally held that there are three types of theories of reference on the table: deflationary theories, description theories, and causal theories.

Deflationary theories can be immediately set aside. They claim *all* that can be said about the nature of reference is the set of true claims generated by instantiating a simple schema, such as:

**Disquotation** ⌜*Tokens of *n* refer, if at all, to n*⌝

Disquotation is from Horwich (1988 p. 119 where the */* marks signal disambiguating quotation). Other deflationary theories are available, though all make use of disquotation in some way (often within the scope of a quantifier over terms or sentences).[83] All such theories are theories of what reference *is*, not of how it is distributed. Comparing his theory with Kripke's causal picture, Horwich is very clear:

> one should distinguish between the facts in virtue of which a word has the meaning it does in a given language [i.e. a theory of what meaning or reference *is*], and the facts in virtue of which a given member of the linguistic community understands the word in that way [a theory of the distribution of meaning or reference]. But then it becomes clear that Kripke's remarks about the inheritance of reference pertain to the second issue, and that they have nothing specifically to do with singular terms.
>
> (Horwich 1988 p. 117; Field 1994 makes similar comments)

Horwich goes on to offer a theory covering the first issue, and ignores the second. But the second issue, not the first, is my concern in this chapter. So deflationary theories, as they have actually been developed, are orthogonal to my concerns.

Moreover, if deflationism *were* offered as a theory of the distribution of reference it would be inapplicable in the present context. It is only possible to

[83] See Båve (2008) for an overview.

state the deflationist thesis for entities which can be quoted and disquoted, such as words and their tokens. But there is no notion of quotation or disquotation for a cognitive token or perceptual object file, nor is there a formal language in which such tokens appear in with the theory could be formulated. Whatever its merits then, deflationism cannot provide a general theory of reference determination for refemes.

The next candidate are descriptivist theories. On a descriptivist theory of reference, any bearer of reference is associated with some description, predicate or property, or set or cluster of these, which is uniquely satisfied by at most one object. This object is the referent. This approach can be applied regardless of the metaphysics of the bearers of reference—thought-episodes, names, utterances, demonstrative acts, etc. could all be associated with a satisfiable content. And I have already claimed that all bearers of content in my theory *are* associated with some satisfiable content. However, as I described earlier, it is a mistake to think that the referent of a refeme must be whatever (if anything) satisfies its ascriptive content. The arguments for this are well-worn, and I gave my own in the previous chapter.

These arguments defeat descriptivism as a theory of *all* reference, but I will return to something like a descriptivist theory with a causal flavour in my discussion of 'primitive reference' below.

So this is the negative argument: deflationary theories are inapplicable and descriptivist theories are false, so something like a causal theory must be the right answer.

Now the positive argument, that this is exactly what we should expect.

Historically causal theories have been called causal theories of *names*, but in practice almost all causal theorists work in terms of the reference of *utterances* and other event-like entities. The general idea is that reference is ultimately to be explained in terms of causal relations between the reference-bearers and their referents.[84]

According to the theory I have given so far, it is exactly these kinds of things which bear reference: utterances, cognitive tokens and object files are all event-like entities within the causal order. Unlike the deflationary theory then, the causal theory accords with my framework, and the theory of refemes I developed in Chapter 3.

While it is true that traditional causal theories of reference don't extend to non-linguistic entities (though see Devitt 1981 esp. p. 133) they do not

[84]See for example Kripke (1980 p. 96), Devitt (1981), Chastain (1975), and Sainsbury (2005). Kaplan's (1990) causal theory of reference is one significant exception. Sainsbury (2014) offers a causal theory of the identity of names over time; he distinguishes this carefully from a causal theory of the reference of those names (pp. 209–10).

explicitly rule them out either—part of the work of this section is to describe how the extension must be made. Conversely, it is true that appeals to causation have been made in theories of the nature of perception and cognition, and often precisely because causation is a clear way of explaining how content is determined (see Grice 1961; Arstila and Pihlainen 2009 on perception; Sullivan 2010 and citations therein on thought).

Finally, the driving insight of the causal theory is that sometimes the reference of some token is determined by, or depends upon, that token's relation with some *other* referring token (this has been called 'borrowing', 'inheritance', 'deference', 'secondary reference', etc.). There are independent reasons to think that reference works this way.

Sitting with a friend in the pub, we overhear the next table talking derisively about someone called 'Louis'. Noticing we are both eavesdropping, I remark 'Louis sounds like a real piece of work'. Surely my utterance of 'Louis' refers to whoever the other table have been talking about. If so, this can only be because of some relation between my utterance, and the other table's previous utterances which I have overheard; my utterance 'inherits' its reference from theirs (this case is a variant of one presented by Evans 1973 p. 192). Or suppose I see a bird on a branch, and continue thinking about it as it flies away out of sight. Surely my initial perceptual reference to the bird is what enables my later cognitive reference to that *same* bird. These cases are particularly stark, but we will see below that many cases of reference can be understood in this way.

These last two points are of particular importance for my view, since I claim that a refeme of one kind (e.g. a name utterance) can inherit its reference from one of another kind (e.g. a visual object file). Since such things are all part of the causal order, this is a straightforward and unmysterious phenomenon.

Just as with ascription though, the chain of inheritance must come to an end somewhere. If some refemes inherit their referential content from others, then somewhere along the way some must refer 'primitively'. We have arrived once again at the primitive/inheritance model of determination.

### 5.3.2 Primitive reference

In this section I describe the conditions under which a refeme refers to something without inheriting its reference from another refeme. The most common and widely discussed cases of primitive reference are perceptual refemes. I spend most of this section developing my theory just for perception (§5.3.2.1), then describe how it can be applied to cognition (§5.3.2.2) and briefly to language (§5.3.2.3). My claim is that a refeme R refers primitively to an object O iff R

is caused by O, O fulfils a significant portion of the properties ascribed by R's ascriptive content, and this ascription fulfilment is dependant upon changes in the object's properties in a way I spell out below. Thus in the case of primitive reference (and only primitive reference), something like the principle of Determination is true, though not without caveats.

I close with two notes on the limitations of primitive reference. First, it is a consequence of my theory that it is impossible to primitively refer to anything from which one is causally isolated. If we are causally isolated in this way from abstracta like $\pi$, or from future existents such as Kaplan's 'Newman1', then we cannot refer to them (§5.3.2.4). Second, I'll claim that it is not always determinate which object some refeme primitively refers to; I'll argue that my theory delivers the right results in all such cases (§5.3.2.5).

#### 5.3.2.1 In perception

The first of my three conditions, causation, will be easily agreed. Consider Grice's (1961) clock case, which I have already discussed at length, in the terms of the framework I have developed. A subject is faced with a clock, and they token a refeme which ascribes all the properties that the clock actually has. But the subject's apparent experience is a drug-induced hallucination—the clock itself has nothing to do with the seeming appearance of a clock. In this case, the refeme does not refer to the clock, though we can easily imagine a similar case in which it does. The difference lies in whether the refeme stands in a relation to the actual clock which it could not if the clock were not there. Whatever this relation is, it will be necessary for primitive reference. I claim the relation is causation. Grice was writing about veridical perception, not reference, but he draws a similar conclusion.

My claim might seem hasty. There is a rich tradition criticising Grice's claim and wider theory (White 1961; Kim 1977; Snowdon 1981; 1992; Snowdon and Robinson 1990). But it is important to distinguish my conclusion from his. Grice was concerned to give an account of perception. My project is much narrower: to give an account of what determines the reference of visual refemes. The clock case is informative on both fronts. It demonstrates that veridical perception (Grice's concern) and primitive reference (my concern) both require some relation between the experience (for perception) or refeme (reference) and the object which could not obtain if the object were not present. Grice assumes that the relation is causation, and his critics have pointed out that there are other possibilities. Snowdon in particular has demonstrated several times that the object (the clock) might partially *constitute* the perceptual experience rather

than causing it, so Grice's conclusion is not immediately demonstrated. But a response like this is only available because Grice is interested in the nature of perceptual experience, and it is not obvious that the relation between object and experience must be causal, or that it must not be constitutive.

But such a response cannot be made against my claim. I claim that for a perceptual refeme to refer to an object, the two must be causally related. The relation between refeme and object cannot be one of constitution. For one, because the refeme is part of the subject's perceptual system, and the object clearly is not. For another, because if the object constituted the refeme, then the relation between them would not do the explanatory work it is supposed to; I am appealing to Grice's clock case to explain something about the nature of reference, the paradigm of a relation between one thing and *another*—not one thing and *its constituent*. But causation is the right kind of relation, and has none of these problems. I claim it is a necessary condition of primitive reference.

It will immediately be agreed that causation is not sufficient. Suppose that unbeknown to me a computer is hooked up to my brain. It alters my visual system such that wherever I would usually see a flock of birds I instead seem to see an helicopter, occupying roughly the same space; if the flock are on the ground I seem to see a grounded helicopter, if they are flying then 'it' is flying. The illusion is perfect: the computer operates the 'helicopter' in such a way as to make it seem real, including compensating for the fact that nobody else seems to see a helicopter. The computer also affects my other senses, screening out inappropriate bird noises and replacing them with rotor/engine noises, making me seem to smell engine fumes instead of bird guano if I'm up close, etc. Things like the size or intensity of certain features translate to the helicopter too: bigger flocks are replaced with bigger helicopters, and I smell more 'fumes' around smellier birds. Finally the helicopter I seem to see bears a unique and distinctive insignia. The computer does nothing else.

In this case, there is a sense in which I am a reliable detector of flocks of birds, and each such detection is causally related to a particular flock. Whenever there is a flock of birds present, my perceptual faculties react in a distinctive way, which correlates reliably with (some of) the flock's characteristics, and not with anything else (note the distinctive insignia—the apparent helicopter does not look like any real helicopter in at least this respect). But I think it is clear that in this case I never perceive any flocks of birds, despite the causal link.

The reason is that the sensory experiences I have are as of a helicopter, making engine sounds and smelling of fumes, though they are causally linked with a flock of birds, making squawking noises and smelling of guano. More generally,

the lesson is that primitive reference requires some substantial level of ascription satisfaction as well as a causal link.

I say 'some substantial level' because it is clear that *complete* satisfaction is not necessary. If my perceptual refeme is caused by an object which satisfies *nearly* all of the refeme's ascriptive content then my refeme still refers to the object. For example, suppose I see a clock, token a visual refeme and ascribe the property *showing the time 10:10*. In fact, the clock's face shows 1:50—by a trick of the light I have confused the hour and minute hand and read the wrong time. So my ascription of the property *showing the time 10:10* is not satisfied. But surely this one failure does not defeat my perceptually referring to the clock—I do refer to the clock, in part because most of the ascriptions I make are correct, and it so happens that one of them is not.

I said in Chapter 4 that a single refeme's ascriptive content can change over time. An object's actual properties might change too. For short-lived refemes like name-utterances, this is irrelevant. But for longer-lived refemes like object files, it is important that the ascriptive content not get out of step with the object's actual properties. So the requirement for substantial satisfaction holds at a time, with respect to the properties the relevant object has *at that time*. Suppose I token an object file caused by a chameleon which is currently green, and ascribe *being green*. If the chameleon turns red but I don't alter my ascription to *being red* instead, surely I no longer see the chameleon.

There are no hard rules about how much of the ascriptive content must be satisfied in order for a refeme to count as referring, or about *which* must be satisfied. This is an advantage of the theory. Suppose I am presented with a hatstand with a coat and hat hanging on it. By a trick of the light it seems to me that I see a person, wearing the very coat and hat which I see. Many of my ascriptions—surely some of the most significant—are false, though many others are true. This doesn't stop me from perceptually referring to the hatstand.[85] The fact that there are no hard rules will also be helpful for dealing with indeterminacy, as we will see below.

A brief aside: this feature frees my account of the 'qua problem' for causal theories of reference. The problem is this: on a purely causal theory of primitive reference determination, a refeme's referent is whatever causes it. I objected to this above (the helicopter/birds case) on the grounds that if the 'wrong kind' of thing is the cause, the refeme doesn't refer. But there is a more subtle problem: even in entirely normal cases, many different things, of different kinds will count as causes of the refeme. Suppose I am confronted with a rabbit, and consequently

[85]Of course, if *all* the ascriptions are false then I simply do not refer to the hatstand or the coat, as in the case of the helicopter/flock of birds.

token a visual refeme. Is my refeme caused by the rabbit, its fur, the visible part of it, the current time-slice or perhaps some combination of these?[86] Without further qualification, any of these might be the 'right kind' of object. Surely they cannot all be the referent! But my account adds a further condition: the referent must satisfy enough of the ascriptions. In the rabbit case, I will likely ascribe *being a rabbit*—this rules out the fur, the visible part, and the time-slice. Other situations may require more ascriptions, but the general solution stands.

An advantage of this approach is that it does not prove too much, as it still possible to refer to these other kinds of thing, simply by changing the ascriptions. If I withhold the ascriptions of *being a rabbit* and make instead *being a fur pelt*, my refeme will refer to the fur, and not the animal.

Returning to the main account: causation and significant satisfaction are both necessary. But they are not jointly sufficient. Consider a different case of computer-altered vision. In this case the computer is mounted on my forehead just above my right eye, so that normally I would see it in my periphery (like seeing the frame of one's glasses). But the computer is hooked up to my brain, and 'superimposes' an image of just such a computer, exactly where the real computer would be visible to me. The superimposition is sensitive to changes in things like my visual attention and environmental light and shade, and reacts accordingly. But it is not sensitive to alterations in the actual computer's appearance: if the casing of the actual computer gets scratched, I will not seem to see a scratched computer.

The resulting experience is phenomenally indistinguishable from the experience I would be having if the computer were there, but were not operating on my visual system (at least it is to begin with—if the computer casing gets scratched this will no longer be true). The two conditions considered so far are both fulfilled: the computer causes my seeming to see a computer (the causal condition), and the computer I seem to see looks just like the computer causing my seeming to see it, modulo some minor possible alterations (the significant satisfactional condition). And yet, it seems that I do not see the computer. So causation and satisfaction together are not sufficient (Dickie 2011b calls these cases of perceptual 'luck').

This case is importantly different from that of the chameleon. There, the concern was that if the chameleon's properties changed and I failed to update my ascriptions in response, then *after* the change I would not count as seeing the chameleon, though I might count as seeing it before, at the time when my ascriptions were satisfied. In the present case though, the concern is that I do

[86]The choice of example is deliberate—the qua problem is related to Quine's (1960) concerns about the inscrutability of reference.

not count as seeing the computer at *any* time, even at the early stage where my ascriptions are satisfied, because of the overall relation between the ascriptions and the object.

I think the problem with this case is that the relation between the object causing the refeme and the ascriptive content it satisfies is of the wrong kind. Intuitively, the success of the ascriptions does not *depend* on the object in the right way. If it did, then I really would see the computer.

We see this by comparing the case with two further cases. First where the computer is mounted on my head but does not affect my vision at all; second where the computer is more sophisticated than above, and alters the superimposed image in line with all environmental changes and alterations to the computer itself. In both of these cases, if the casing of the actual computer gets scratched, I *will* seem to see a scratched computer. Clearly I perceptually refer to the computer in the first new case. I think I also see it in the second—the situation is similar to a viewing a remote object via a live video feed, which we would naturally take to be a case of perceptual reference.

The important thing for my argument here is the difference between these cases and the case of the less sophisticated computer. In all three cases, what I seem to see depends to some extent upon what is actually there. I seem to see *something* because in every case there is something there, which causes my seeming to see it. Likewise in every case I seem to see a computer of a certain shape and colour, etc., and the thing which is causing my visual refeme *is* a computer of that shape and colour. So both the causation and significant satisfaction conditions are fulfilled. The difference between the first case and the later two is that in the later two, a far greater proportion of my satisfied ascriptions are counterfactually dependant on the properties which satisfy them; they are *reliable* and not *lucky*.

When the computer doesn't affect my vision at all, if the computer's casing gets scratched, I'll seem to see a scratched computer because I do see a scratched computer in just the normal way. When the computer manipulates my vision but updates on all possible changes, if the casing gets scratched, again I'll seem to see a scratched computer, because the computer I'm seeing has been scratched (though this information arrives by a more convoluted route). In the first case though, this doesn't happen—a large proportion of my ascriptions are true, but wouldn't remain so under small changes. The problem is that the ascriptions will not 'update' in line with changes in the world.

I claim that this is the third condition on primitive reference. For a refeme to get its reference fixed primitively to an object, the object must cause the refeme, a significant portion of the refeme's ascriptive content must be satisfied by the

object, and the agent's ascriptions must change and update roughly in step with the features of the object which make them satisfied.

More formally, my third claim is that for a significant proportion of the refeme's ascriptions, the agent producing the refeme must have dispositions such that: if the object's features change such that, were the refeme's ascriptive content unaltered, that ascription would no longer be satisfied, then the agent will either withdraw or alter the ascription. Or in other words: for a significant portion of their ascriptions, the agent is disposed to update whether and how they make that ascription following changes in the object's properties upon which that ascription's satisfaction depends.

We now see why my perceptual refeme does not refer to the computer in the first case, but does in the second two. In the first case, although the computer itself causes my refeme, and the bulk of the refeme's ascriptive content is satisfied by the computer, I am not disposed to alter or withdraw any of my ascriptions following changes in the computer's properties. The satisfaction of my ascriptions is *fragile* under changes in the properties which satisfy them. Not so in the other two cases. When the computer is present but does not affect me, I will update my ascriptions in just the normal way. When the computer is present, affects me *and* is sensitive to all the ways things could change, the situation is indistinguishable (for me) from one where the computer does not affect me, so again I will update in the normal way. In both these cases, the satisfaction of my ascriptions is *robust* under changes in the properties which satisfy them.

Two clarifications on this third condition. First, like the condition of significant satisfaction, there is no requirement that *all* the refeme's ascriptions are counterfactually dependant in this way, just that a significant enough proportion are. And this is important for similar reasons: if just one or two of my ascriptions fail to be counterfactually dependant in the right way, that doesn't defeat my referring to the computer.

Second, the condition is spelled out in terms of *dispositions*, but these dispositions need not ever be actualised. It is irrelevant whether I actually *do* ever alter the ascriptive contents of my refeme. What is required for primitive reference fixing is that the agent has the *disposition* to withdraw or alter the relevant ascriptions under changes of circumstances, not that they actually do this. Without this caveat, it would be impossible to perceptually refer to an object which did not visibly change while one was looking at it!

#### 5.3.2.2 In cognition

We now turn to primitive reference in cognition. Cognitive refemes are tokens of concepts. Since I've already sketched reasons for holding the view I put forward, I'll just describe a plausible situation involving a cognitive refeme with primitively-fixed reference and show that my theory correctly counts it as such. I'll also show that any alteration to the case which makes it fail one of my three conditions will also alter our intuitions about whether it is a case of reference—thus that my theory accurately captures all and only primitive cognitive reference.

Consider the detective in a very small, quiet town where several residents report their dogs missing, all of them purebred pedigrees. Unaware of each other, none of the residents suspects foul play. The detective has all the facts though and seeing a connection she realises that there must be a single man stealing the dogs, and that he must be tall, knowledgeable about dogs, and new to the town.

As it happens, the detective is almost entirely correct. There *is* a single person stealing the dogs, and he is a tall man, new to the town. He knows nothing about dogs though—he has been taking them at random, and it is only by chance that all the dogs so far are pedigrees. Let us call the criminal 'X'. The detective tokens a cognitive refeme in her thought of the criminal, with ascriptive contents *tall*, *knowledgeable about dogs*, and *new to the town*, and she refers to X.

This is a case of primitively-fixed cognitive reference. The detective has never encountered X, and since none of the residents suspects a criminal, none of them has thought of him in any way, so her reference cannot inherit from theirs. Nevertheless she refers.

This is because the cognitive token she produces fulfils all three conditions above. It is caused by X: if he had not been around and stolen all the dogs, then she would not have produced it. A significant portion of its ascriptive content is satisfied by X: he is a man, is tall, and is new to the area. Though he is not knowledgeable about dogs, this alone doesn't defeat the detective's referring. Finally, the detective is disposed to alter or withdraw the ascriptions she makes in line with alterations in the referent. Suppose that during the existence of the detective's initial cognitive refeme, someone radios about a tall man fleeing town on a motorbike and a dog in the sidecar. Realising that the rider is X, the detective will withdraw her ascription of *new to town*, since he is no longer around. The fact that the detective is disposed to make such alterations—even before the news actually comes through—secures her reference. So the three conditions together are sufficient for primitively-fixed cognitive reference.

Each condition is also necessary. In versions of the case altered so that one of the conditions fails, the detective doesn't refer to X. First, suppose X doesn't cause the detective's refeme. She does learn about the lost dogs, but doesn't piece the evidence together. Instead, she happens to watch a detective drama about a tall man who kidnaps dogs, and tokens a refeme caused by that. By chance, most of the refeme's ascriptive content is satisfied by X. But still it is clear that she doesn't refer to X.[87] Second, suppose the detective's refeme is caused by X, but he satisfies little or none of its ascriptive content. The detective might token a refeme and ascribe *being a dog*, *being a pedigree breed*, and *being recently stolen*. She does so because there is something of which all this is true, and that is only the case because X has stolen a dog. So X is in the refeme's causal history, though does not satisfy a significant proportion of its ascriptions. We may even stipulate that—bizarrely—the detective is disposed to revise the ascriptions she makes on the basis of observed changes in X (even though X fulfils none of them at the moment). Clearly she doesn't refer in this case either. Third, suppose that X causes the detective's refeme, and a significant portion of its ascriptive content is satisfied, but she is not disposed to update these ascriptions when X's properties change. In my phrase above, the satisfaction is 'fragile', in Dickie's it is 'lucky'. Here too, the detective does not refer to X. This should be intuitive if we consider what would happen if X had been arrested and brought into the room just after the detective works out he must exist—the detective's initial primitively referring refeme will continue to exist, but even when she learns that the man in front of her is X (i.e. is the man she is thinking of) and that he doesn't know much about dogs, she won't update her ascriptions (because she lacks the relevant dispositions). This case is unlike the others in that the problem is something to do with the detective herself—she has not mastered the use of concepts to relate to worldly objects. Nonetheless, as with the other two conditions, this condition's failure entails that she doesn't refer.

The success of my theory in describing such cases, and that it need not be altered as it is transposed from perception to cognition, I take as evidence for it.

#### 5.3.2.3 In language

If it happens at all, it is comparatively rare that linguistic refemes have their reference fixed primitively. The principle linguistic concern of this thesis has been utterances of names. These never have their reference primitively fixed. Names are bestowed on objects, and then used later, but the initial referring

[87] It is not obvious to me whether the third condition, of being disposed to revise her ascriptions, is also met in this case. But even if a similar case could be constructed in which it definitely was, I am confident that the detective would not then count as referring.

element in the chain is never an utterance of the name bestowed, it is another refeme, from which later name-utterances inherit their reference.

The act of bestowing a name on an object is called a 'baptism' or 'dubbing'. Though baptisms play a significant role in causal theories of reference for names, surprisingly little has been written on their nature or mechanics.[88]

Whatever else is to be said about them though, baptisms clearly do not constitute or include a way of fixing the reference of any particular utterance, besides primitive and inherited reference. In a baptism, the object is referred to by some device (an existing name, or demonstrative, etc.) which is not peculiar to baptisms, and then a name is specified, and the two are somehow related such that later uses of the name refer to the object.[89] But none of this requires any special kind of reference. Baptisms are likely a particular, specific way of providing links in the chain of inheritance, but clearly baptisms don't provide a kind of reference which is *neither* primitive nor inherited. I won't provide a theory of the nature of baptisms here.

All that said, there may be cases of primitively fixed linguistic reference; they are just not utterances of names. I offer here some very speculative comments.

Demonstratives might seem a reasonable candidate. Roughly, a demonstrative is an utterance of an object-indicating term which is sensitive to context ('this' or 'that') coupled with a physical indication of an object by pointing, nodding or gesturing. The literature is vast, but all agree that a demonstrative generally includes these features (though not every instance of a demonstrative includes every one—in the right context one can demonstrate with only a word, or a nod). Demonstratives are referential; they are used to refer to objects, and *which* object a demonstrative refers to is determined by features of the utterance, the gesture, and possibly further factors like the speaker's intentions. There are many theories on how exactly this happens, but they can be helpfully divided in two. Views of the first kind say that in the whole procedure, one element really picks out an object, and the demonstrative as a whole just refers to that (even if the demonstrative includes other features). For example, it might be that a demonstrative refers to whatever object the speaker intends to refer to, regardless of gesturing or speaking (Donnellan 1966; Kaplan 1989b; Bertolet 1980; 1986). Or perhaps it refers to whichever they perceptually focus on (Kaplan 1989a), or whichever they point at (Reimer 1991), etc. On any such view,

[88]For some rather scattered comments in the tradition of the causal theory, see Kripke (1980 p. 96), Evans (1973), Sainsbury (2005 Section 3.5), deRosset (2020), Michaelson (2023a), and Assadian and Sbardolini (2023). Austin (1965) also includes careful discussion of naming acts (e.g. p. 5) though his interest is in their status as linguistic actions, not especially in names. See Heagren (2026) for a developed view.

[89]We might say that the reference of a *name* is fixed or determined this way, but this is a very different notion from that I am concerned with here.

reference is primarily secured by this single element, which is distinct from the demonstrative as a whole. The intentions, perception or pointing may be *part* of the overall demonstrative, or may partly *constitute* it (or relate to it in some other way) but they are not identical with it, and the demonstrative itself just refers to whatever this distinct element refers to. So the demonstrative refers to its referent because it is related to something distinct, which itself refers to that referent. Or in other words, according to these theories, demonstratives inherit their reference. Views of this kind do not support demonstratives as cases of primitively fixed linguistic reference.

Views of the second kind are more promising. They say that no one element of the demonstrative refers to anything on its own, but that each contributes a condition or property which the referent, if any, must satisfy, and the total demonstrative refers to whatever satisfies them all. For example, Reimer (1992) says that a demonstrative refers to whichever object is both in the path along the speaker's finger's pointing and is a member of the kind of object they had intended to demonstrate (regardless of whether it *is* the object they intended to demonstrate). Each condition alone underdetermines reference, but when one successfully demonstrates there is only one thing which satisfies both. Contextualist views like Wettstein's (1984) could be read in a similar way. On such views, it really is the *demonstrative* which primitively refers, not any of its components.

On the latter view demonstratives come out somewhat like definite descriptions, in that if they refer to anything, it is whatever thing fulfils certain conditions, determined by the features of the demonstrative. For similar reasons, utterances of definite descriptions are good candidates for primitively fixed linguistic reference. An utterance of a definite description composed and spoken 'on the fly' without prior cognition, might form a linguistic analogue to the cognitive tokens of the detective in §5.3.2.2. Such an utterance is made up of several smaller utterances of kind terms (e.g. 'King', 'bald', etc.) and of logical or grammatical terms ('the', 'not', 'if', 'is', etc.). Plausibly, each utterance of a kind term will come with certain ascriptive content (such that 'King' ascribes *being a king*), and there may be rules of composition relating these ascriptive contents together, accounting for the roles of the logical and grammatical terms. Taken as a whole then, the utterance of an entire definite description could have a single ascriptive content, determined by those of its parts and their relations. This ascriptive content would then play the standard role in determining the reference, if any, of the utterance of the description.

This is a plausible sketch of how primitive reference might occur in language. The model of demonstratives or of definite descriptions could be developed in

more detail into a full account in line with the rest of this thesis. That is not the project here though.

#### 5.3.2.4 Apparent reference in spite of causal isolation

Several authors have discussed apparent cases of reference to things from which we are entirely causally isolated, presenting these as problems for causal theories of reference. There are two kinds of case.

First, it is suggested that we can dub entities which do not yet exist at the time of the dubbing, and thus refer to them before they exist. Kaplan (1968 p. 201) suggests announcing 'I hereby dub the first child to be born in the twenty-second century 'Newman 1" and then using the name. Jeshion (2000 p. 300) imagines a doctor selecting a particular sperm Sz and egg Ez and announcing "Zeke' is to refer to the product of the union of the sperm Sz and [egg] Ez'. Sainsbury (2005 p. 109) and Hawthorne and Manley (2012 pp. 27–28) discuss similar cases.

Second, it is noted that we seem to comfortably use names for abstracta, like numbers and sets. It is perfectly meaningful to say that $\pi$ is an irrational number, or that the empty set is a primitive in maths. Making sense of this has been a concern for several causal theories. (e.g. Vendler 1976 Section 7; Kripke 1980 p. 60 on $\pi$; Searle 1983 p. 235)

The problem in each case is the same. I have claimed that dubbing requires reference to the thing dubbed, that reference is either primitive or inherited, that the latter requires the former, and that the former requires causation. So it is a consequence of my theory that it is only possible to refer to future entities or abstracta if it is possible that they cause a dubbing. Conversely it is impossible to dub or refer to anything we are causally isolated from. This is a problem because it seems that we are causally isolated from future entities and abstracta, and yet also that we meaningfully use names for them.[90]

The problem is neither original nor pressing; really it is just an old problem in disguise. Notice that the objection comes in two steps: first that on a causalist view of reference determination, it will be impossible to refer to future particulars or abstracta. This is clearly true, and is the only part which is really relevant to the topic of this chapter. Second that therefore it will be impossible to account for the content of refemes which apparently refer to future particulars or abstracta. But this is just another way of saying that these utterances will not refer at all, and that is worrying because they seem to be meaningful. So the problem of causal isolation is really just a case of the problem of intelligible empty names.

[90]Or so it is argued; see e.g. Linnebo (2012) for an attempted defence of reference to abstracta compatible with broadly causal intuitions, roughly by rejecting the idea that we are entirely causally isolated from abstracta.

The solution is the same as before: even if utterances supposed to refer to future particulars or abstracta really don't refer, this does not make our behaviour surrounding them unintelligible, because their content is not exhausted by their (lack of) referents. Each utterance will also have ascriptive content and (as with other empty utterances) this will serve to explain its apparent meaningfulness.

A more sophisticated objection might now arise. According to my theory, it is impossible to refer to future particulars or abstracta. But if the theory is true, it is impossible to discuss these things at all, even to say of them that they are beyond our reference. And if this is impossible, then by its own lights the theory cannot be truly stated!

Not so. To state my claims about future particulars and abstracta, it is necessary to *quantify over* them, not necessarily to refer to any of them individually. I have not offered a theory of quantification, and nothing in my theory so far suggests that quantification will require the same things reference does (i.e. causation). Indeed, it seems very likely that we can quantify over things to which we cannot refer—it is perfectly sensible to say that there are no unicorns. My claim about abstracta comes out roughly as the claim that for anything which is an abstractum, and for any utterance of a name, the utterance does not refer to the abstractum, because they cannot be causally related (and similarly for future particulars). This can be true regardless of the nature of quantification, and regardless of whether there are any future particulars or abstracta at all.

#### 5.3.2.5 Indeterminacy

Reference and its primitive fixation are sometimes indeterminate. In this section I defend this general idea, discuss two cases, and show that my theory makes good predictions in each of three ways of reading them.

Let us begin with a simple motivational case, exhibiting two different kinds of indeterminacy.

In the first case, when walking in the forest on a dark night, I encounter what seems to be a tree, swaying in the wind. I token a visual refeme, caused by the object, and ascribe properties such as *being a tree*, *moving*, *being three metres tall*, etc. In fact it is not a tree at all, but two people, one on the other's shoulders, each holding a tree branch in each hand, engaged in a ritual.

It is not clear whether I perceptually refer to the people. On one hand, I seem to visually encounter an object, and there is just such an object, which I did encounter. On the other, I have so completely misapprehended it that it is doubtful whether I really refer to *it*, if I refer to anything at all.

In the second case, I stand in a field, watching the sunset behind a stand of trees which separates this field from the next. My eyes rest on the silhouette of a dead tree. Its trunk is slim, and most of the bulk of the image is made of up its many branches, going off at crazy angles. I token a visual refeme, caused by the tree, which ascribes *having many branches*, *being a tree*, etc.. All of the ascriptive content is satisfied by the tree. As I watch, the wind catches the tree, and a large proportion of the branches fall off. Because my eyesight is poor, and many of the surrounding trees' branches overlap with the dead one's, I take this to be a general movement of the branches in the breeze and do not revise my ascription of *having many branches*, though in fact it is no longer satisfied. Thus, before the wind caught the tree, I lacked the appropriate disposition to revise my ascriptions in light of changes in the relevant object. It is thus at least questionable whether my visual refeme really refers to the tree.

My first claim is that if it is indeterminate whether a sufficient proportion of a refeme's ascriptive contents are satisfied by an object for it to refer to that object (and the other conditions for reference are satisfied) then it is indeterminate whether the refeme refers to the object. So in the first case, it is indeterminate whether my refeme refers to the two people because it is indeterminate whether a sufficient proportion of my refeme's ascriptive contents are satisfied by the people in order for it to refer to them. I take this to recover an important intuition.

The indeterminacy in the second case could be read in two ways, covered by my second and third claims.

My second claim is that if it is indeterminate whether enough of the ascriptions are such that the agent is disposed to revise them following changes in the object, then it is indeterminate whether the refeme refers to that object. One way of reading the second case is that I am not disposed to revise some of my ascriptions in the right way under relevant changes, and therefore I don't really refer to it.

My third claim is that if it is indeterminate whether enough of the ascriptions are such that the agent is disposed to *sufficiently* revise them following changes in the object, then it is indeterminate whether the refeme refers to that object. On a slightly different reading of the second case, I *do* have revisionary dispositions for all of the ascriptions I make, but those dispositions do not apply to enough of the ways that the object could change to 'count'. The case was deliberately constructed to include a relatively minor change in the way the object looks; plausibly, I am disposed to revise my ascription of *having many branches* if the tree visibly topples and sheds its branches on the way, but not just if the branches are knocked off by the wind. In this kind of case, I do have the relevant dispositions, but it is indeterminate whether they play the right role (or play it 'enough'), and

therefore indeterminate whether I refer to the tree. Once again, I take this to recover an important intuition.

### 5.3.3 Inherited reference

In this section I describe the conditions under which one refeme inherits its reference from another. Until this point I have treated causal theories of reference as coterminous with primitive/inherited theories of reference determination. According to a crude version of the causal theory, one token inherits its reference from another just when the latter causes the former. My theory will depart from this crude picture in three ways.

First, that causation is not necessary for inheritance. My theory replaces causation with the more general notion of an 'I-relation' (§5.3.3.2). I-relations are necessary for inheritance, and causation is one such, but there are others besides.

I-relations alone are not sufficient for reference inheritance. My second claim is that as well this, the agent producing the second refeme must be be disposed to judge that both are of the same kind (§5.3.3.2), and my third that the ascriptions of the two tokens must significantly overlap (§5.3.3.3).

As with primitive reference, I close this section with some comments on indeterminacy (§5.3.3.4). It is not always determinate which object some refeme refers to; I'll argue that my theory delivers the right results in all such cases.

The account I present has none of the appealing neat simplicity of Kripke's 'picture', and though I argue for my claims, I have not attempted to demonstrate that no alternative is viable. Rather it is a conjecture, presented as a best effort to account for the available data, including the many existing puzzles of reference, and some new ones besides. These data are complex and messy, as many others have noticed, and my account reflects this. I have also tried to present more explicit and detailed answers than many before me (especially Evans and Sainsbury). I hope that my theory is thus more open to careful scrutiny, and presents a more illuminating view, even if this comes at the price of complexity.

#### 5.3.3.1 Causation and refeme systems

Causation is not necessary for the inheritance of reference, even in cases where it is clear that reference *is* inherited.

Suppose I set out to write a history of women in recent philosophy of language. I start with 'I will use the name 'R' for Ruth Millikan'. Later in the book, I write 'R had profound insights—she developed the New Theory of Reference'. This later token of 'R' inherits its reference (i.e. Millikan) from the earlier sentence.

Later I realise my mistake—it was Ruth *Barcan Marcus* who originated the New Theory of Reference. I go back and edit the earlier part to 'I will use the name 'R' for Ruth Barcan Marcus'.

The case is illustrated in the following diagram. I use square brackets with subscripts to indicate names by which I refer to different tokens later in this section.

| Time | Edit | Resulting text |
|---|---|---|
| T1 | I will use the term 'R' for Ruth Millikan. | I will use the term 'R' for [Ruth Millikan]$_{I1}$. |
| T2 | R had profound insights—she invented the new theory of reference. | I will use the term 'R' for [Ruth Millikan]$_{I1}$. … [R]$_{I2}$ had profound insights—she invented the new theory of reference. |
| T3 | (remove original token of 'Millikan', replace with) Barcan Marcus. | I will use the term 'R' for [Ruth Barcan Marcus]$_{I3}$. … [R]$_{I2}$ had profound insights—she invented the new theory of reference. |

After the edit, the latter part of the book has different semantics from what it had before the edit—at T2, I2 referred to Millikan, and at T3 it refers to Marcus. But I2 was produced *before* the edit was made in T3. Clearly the (edited) inscription near the beginning plays a part in determining the semantics of the inscription near the end; I2 'inherits' its reference by a relation with I3. But since the edit (and the resulting token I3) did not cause the token I2, the relation of semantic determination cannot be causal. It seems clear to me that the inheritance relation here is something more like a grammatical or anaphoric relation.

This is not to say that reference is *never* inherited along a causal chain. The token of 'Ruth Barcan Marcus' at the beginning of the book, (though produced at the end of the case), refers to Ruth Barcan Marcus. Clearly it does not get its reference primitively, nor does it inherit along a grammatical chain. This seems like a very standard case of causal reference inheritance.

So some tokens inherit along causal lines and some along other lines.

The causal theory as generally understood adopts the primitive/inherited structure I describe above, and goes on to claim that the inheritance relation coincides with causation. In light of the above case, I will claim that there is no *one* relation which coincides with reference inheritance, but that different kinds of token inherit in different ways.

A *reference-system* is a set of refemes, with a transitive, non-reflexive, asymmetric relation over the set, under which the set is closed. Call such a relation an I-relation, for 'Inheritance'.

For example, the book I imagined writing in the case above is a reference-system. The set of refemes will be the set of all inscriptions of names in the book, and the I-relation will be the grammatical or anaphoric relation I alluded to. This relation relates inscriptions to other inscriptions in the same book, but not to any other inscriptions or refemes. Since all the inscriptions in the book are members of the set, the set is closed under the relation. Call this reference-system 'B'.

A single refeme might be part of more than one reference-system. The token of 'Ruth Barcan Marcus' I eventually produce at the beginning of the book is a member of B. But it is also a member of a much larger reference-system, including all of the utterances and inscriptions of names in its causal history (and all those causally related to them, and so on). *This* system's I-relation is causation. Call this system 'C'.

For any refeme r in a reference-system X, let us call the set of all refemes in X to which r is related by X's I-relation, r's *X-ancestors*. So the token of 'Ruth Barcan Marcus' has no B-ancestors, and many C-ancestors. A refeme's *ancestors* (unqualified) are just its ancestors in any reference-system.

Finally, it will be helpful to have the notion of a *longest terminal chain*. For a refeme r in a reference-system $\phi$, a terminal chain from r in $\phi$ is a finite non-empty series of refemes beginning with r, each of which is I-related, by $\phi$'s I-relation, with the next refeme in the series. In other words, a terminal chain is a finite line of inheritance through r's ancestors in a single reference-system. For example, if the relation is causation, then a terminal chain from r will consist of r, a refeme r' which is a cause of r, a refeme r'' which is a cause of r', and so on, until some last refeme r$*$ (r$*$ might have a cause, but not a cause which is a refeme; it might also be related to other refemes, but not by causation). A refeme with no ancestors in a reference-system has no terminal chain in that system.

Some terminal chains are longer than others. Since causation is transitive, if there is a (perhaps very long) chain of causation linking r with r', then r'', and so on until r$*$, there will also be a (much shorter) chain linking r with 'directly' with r$*$. The former chain is longer than the latter.

Now, a general claim about reference inheritance. When a refeme r is a member of two reference-systems $\phi$, $\psi$ (and perhaps others), such that it has both $\phi$-ancestors and $\psi$-ancestors (and perhaps others), r inherits its reference (if any) along the I-relation associated with the reference-system in which r's longest terminal chain is shortest. So if r is in just $\phi$ and $\psi$, and r's longest terminal chain in $\phi$ is shorter than its longest terminal chain in $\psi$, then r inherits

its reference (if any) along $\phi$'s I-relation, not $\psi$'s. That is, r will co-refer with some refeme in $\phi$, to which it is related by $\phi$'s I-relation. When r is a member of only one reference-system, it inherits along that system's I-relation. An intuitive interpretation of this claim is that reference is inherited along more 'specific' relations when available.

To see how this works in practice, my claim can be applied to standard cases of spoken names, and to the more complex book case above.

In the case of a spoken name, all the name-utterances which are causally related to each other will form a large reference-system. Each utterance will only be a member of that system, so the relevant longest terminal chain will be the chain in that system, leading from the utterance eventually back to an object. The utterance refers to that object. Say for example that I token an utterance of 'Saul Kripke'. The reference-system is the set of all causally related utterances, its I-relation is causation, and the longest terminal chain will be one leading back to Kripke himself. Thus I refer to Kripke.

In the book case, the book I write constitutes a reference-system: there is a set of refemes—the inscriptions in the book—and they are related by a grammatical or anaphoric relation under which the set is closed. This is reference-system B. Since all the refemes are part of the causal order, they will also all be members of reference-system C. At T1 there is only one refeme in B: I1. I1 has no B-ancestors, and therefore no terminal chain in B. It does have C-ancestors though, and thus I1 inherits its reference from a refeme to which it is causally related, since this is C's I-relation. There is some refeme way back in the past which refers to Ruth Millikan, and which causes I1, and in virtue of this causal relationship, I1 refers to Millikan. (The story is not quite this simple, as we will see below, but this is enough for the discussion at hand.)

At T2, I2 is introduced. I2 has both B-ancestors and C-ancestors. Its longest terminal chain in B will reach back through the book, eventually to I1, whereas its longest terminal chain in C will reach way back in history. Since the B chain is the shorter one, I2 inherits its reference (if any) along B's I-relation—a grammatical or anaphoric relation which terminates in I1. I1 refers to Millikan, and at T2 I2 inherits from I1, so at T2 I2 refers to Millikan as well. It so happens that I1 is also among I2's C-ancestors, but this is irrelevant to determining I2's reference.

At T3, I1 is destroyed and replaced with I3. I3 has no B-ancestors, and therefore no terminal chain in B. It does have C-ancestors though, and thus I3 inherits its reference from a refeme to which it is causally related. There is some refeme way back in the past which refers to Ruth Barcan Marcus, and which causes I3, and in virtue of this causal relationship, I3 refers to Barcan Marcus. Since I1 no longer exists, I2 no longer stands in the relevant grammatical relation

with it, so does not inherit its reference, and so does not refer to Millikan. At T3, the shortest among the longest terminal chains for I2 is its chain in B, but at T3 this terminates with I3. I3 refers to Barcan Marcus, and at T3 I2 inherits from I3, so at T3 I2 refers to Barcan Marcus as well.

My theory thus delivers the correct results for this case: the tokens near the beginning of the book each inherit their reference casually, the one near the end inherits grammatically, and its reference changes over the course of the case, as the tokens near the beginning are altered.

The notion of I-relation forms the basis of my theory of reference inheritance.

#### 5.3.3.2 Taking tokens to be of the same kind

Being I-related is not sufficient to determine that one token inherits its reference from another. Suppose I say 'Alice is married to Bob. And Bob is a mechanic'. There are three refemes: the token of 'Alice' (call this U1), the first token of 'Bob' (U2), and the last token of 'Bob' (U3). U1, U2 and U3 all belong to reference system C, with causation as its I-relation. U1 and U2 are both in the causal history of U3. Though U3 (token of 'Bob') might well inherit its reference from U2 (also 'Bob'), surely it does not inherit its reference from U1 (token of 'Alice'). So although all three are causally related, U3 does inherit from one, and does not inherit from the other.

There are cognitive cases of a similar structure. I token a cognitive refeme referring to Alice (C1), thinking of Alice makes me think of Bob (C2), then I forget about it, and later I recall that I thought of Bob earlier and think of him again (C3). C3 has both C1 and C2 in its causal history. Though C3 might well inherit its reference from C2 , surely it does not inherit its reference from C1.

There are even inter-modal cases. Someone tells me of one 'Deleuze', a great philosopher of whom I have never heard before. I ask them to note down a good book of his and they produce an inscription of 'Bergsonism'. Later on in the bookshop I read the note aloud and utter 'Bergsonism' asking for a copy. The first utterance (of 'Deleuze') causes the inscription ('Bergsonism'), which causes the second utterance (also 'Bergsonism'). The final utterance has both the inscription and the early utterance in its causal history. Though the final utterance might well inherit its reference from the inscription, surely it does not inherit its reference from the early utterance.

So even if I-being related is necessary for reference inheritance, it is not sufficient. There must be some further condition, specifying which refemes inherit from each other and which do not.

This issue has seen little discussion in the existing literature on causal theories. When it is mentioned, it is assumed that it sufficient to stipulate that causally related utterances must be of the same name. At first pass this solution might seem obvious: utterances of 'Bob' can inherit from other utterances of 'Bob', because they are utterances of the *same name*; though utterances of 'Alice' are not. This straightforwardly accounts for the initial linguistic case, and the inter-modal case of 'Deleuze'. A similar claim could be formulated for cognition in terms of tokens of the same concept and the idea could be extended to perception as well.

There is clearly something to this idea, but it has three problems as it stands. First, it only applies to inheritance between refemes of the same kind. There is no case in which a cognitive token and an utterance are utterances of the same name, or tokens of the same concept. But it is an important claim of my theory that reference can be inherited between refemes of different domains—perceptual to cognitive, cognitive to linguistic, etc. Second, it is not clear how to individuate kinds such that utterances of 'Bob' (U2 and U3) count as the same kind, but those of 'Alice' (U1) do not. There will be some ways which wrongly count all such utterances together: all are utterances, all are utterances of names, all are intelligible speech acts, etc.. Conversely there will be some which wrongly distinguish U2 from U3 though both are utterances of 'Bob': they might sound different, or be uttered at different volumes. Abbreviations, speech impediments, and differences of formality ('sir', use of initials, last names etc.) will further complicate the picture.[91] Third and compounding the second, the conditions for counting as the same kind vary with context. Some languages (Mandarin) are much stricter than others (English) in whether tonally different utterances count as the same. Even stipulating the *same* Mandarin name, in an English conversation between a native Mandarin speaker (with poor English) and a native English speaker (with poor Mandarin), the threshold for counting as pronouncing the name will be lower than it would in a Mandarin conversation between two competent native speakers. Before widespread standardisation, the same name (of the same person) might be spelled differently in different inscriptions; today at least one of these would be counted wrong.

There was clearly something in the claim that only similar tokens can inherit from one another. But in view of all this, it is fruitless to find criteria for when two tokens are 'objectively' similar—the conditions are too flexible. Instead I claim, the relevant criterion is subjective: it should be defined in terms of whether a certain agent, in a certain context, judges them to be similar. I give

[91]On the vagaries of these and other, related concerns, see among others Kaplan (1990), Sainsbury (2014), Gray (2015), and Michaelson (2023b).

more details below, but the principle strength of this account is that there is no appeal to 'actual' similarity, so the problems above do not arise.

An agent-specific account has its own problems though. People do not generally go around making judgements about the kind-similarity of different refemes (not even 'tacit' judgements). One might appeal to an idealised agent, but it is very difficult to specify exactly how this idealisation is supposed to go (compare Nowak and Michaelson 2020).

So I claim, the similarity condition must be described in terms of a particular agent, at a particular time. The agent in question must be one who could be aware of both refemes, so the obvious candidate is whichever agent produces the second refeme. I claim that in order for one token to inherit from another, it is necessary that the producer of the latter token be disposed, at the time of its production, to judge them to be of a common kind if asked, a kind which could support them co-referring. This solves both issues with a subjective account. Appealing to a particular agent avoids the problems of specifying an appropriate idealisation. Appealing to dispositions means that the agent need not actually *make* any judgements.

Such dispositions to judge kind similarity can bridge the divide between domains. An agent can token a perceptual refeme, then later a cognitive refeme, and be disposed to judge that they are of the same kind. Importantly, it is irrelevant whether they are *actually* members of the same kind. Plausibly metaphysical investigation would reveal that they are not. Nevertheless, an agent operating with these refemes may be disposed to *judge* that they are of a similar kind (few agents are philosophically sophisticated, and even those who are rarely apply their principles in each everyday action). Of course in practice, agents will rarely have a disposition to judge utterances of 'Alice' and 'Bob' as being of the same kind, so intuitive expectations about such cases are recovered.

Appealing to a particular agent's dispositions, at the particular time they produce a certain refeme, also solves the second and third problems: in that particular context, that particular agent either would or would not judge two refemes to be of a similar kind. The question of what standards similarity should generally be judged by falls away entirely. Speakers in a Mandarin conversation will hold stricter dispositions to judge kind similarity than those in English conversations, so the intuitive expectations are recovered here too.

Let us see how this works in practice. In the first case I say 'Alice is married to Bob. And Bob is a mechanic' uttering U1, U2 and U3. The question was why U3 inherits its reference from U2, but not U1, when all three utterances are united by the same I-relation. The answer is that I, the producer of U3, would judge U2

and U3 to be of a single kind (utterances of the same name), but would not so judge of U1 and U3 (or U1 and U2).

The cognitive case works the same way. I am aware that my cognitive token referring to Alice (C1) is very different from my two later cognitive tokens referring to Bob (C2 and C3).[92] So if it came to it, I would judge the two later tokens to be of a kind, but not the earlier token. Thus C3 can inherit its reference from C2, but not from C1.

Finally in the inter-modal case, I judge the inscription of 'Bergsonism' to be of a kind with my utterance of 'Bergsonism', but not the earlier utterance of 'Deleuze'.

#### 5.3.3.3 Ascription overlap

We have seen so far that for one token to inherit its reference from another requires at least that they are related by a relevant I-relation (causation is one such relation, though not the only one) and that they would be judged to be relevantly similar by the producer of the latter refeme (latter by that I-relation). But fulfilling these two conditions is not sufficient to determine that one inherits its reference from the other. Consider Evans' (1973) famous case of 'Madagascar'. Evans quotes from I. Taylor (1896):

> In the case of 'Madagascar' a hearsay report of Malay or Arab sailors misunderstood by Marco Polo … has had the effect of transferring a corrupt form of the name of a portion of the African mainland to the great African Island.
>
> (Evans 1973 p. 196)

To redescribe the case as it is relevant here: originally, the name 'Madagascar' was used only by African locals to refer to an area of the African mainland. Let us call one of these utterances U1. Some foreign sailors arrive in southern Africa, pick up the name from the locals, and go on using it to discuss this mainland area among themselves after leaving. The sailors then travel to Asia, where they meet Marco Polo. He hears them use the name, mistakenly comes to believe that they refer to an island off the south east coast of Africa, and begins to use it this way himself. Via Polo's subsequent writings, this usage spreads throughout Europe, and eventually to us in the present day. It is widely agreed that our uses

[92] What this awareness consists in is not the concern here. It is an advantage of my account that we side-step questions about whether two cognitive tokens really *are* of a type, and thus of what these cognitive types might be, or how an agent would be aware of them.

refer to the island, not the mainland. Call one of these uses U2, and another, subsequent use U3, such that U2 is among the causes of U3.[93]

U1, U2 and U3 all belong to a reference system with causation as its I-relation. They are all part of the same causal chain. And U1, U2 and U3 would all be judged to be similar by any of their respective producers.[94] If being I-related and such as to be judged similar was sufficient for inheriting reference, then U3 would co-refer with U2 (and U1) and U2 would co-refer with U1. But U3 and U2 refer to the island, and U1 to the mainland. Being I-related and judged similar are both necessary, but not jointly sufficient, for inheriting reference.

The third condition for one refeme to inherit its reference from another is that the ascriptive content of both refemes must overlap significantly. The ascriptive content need not be identical, and there are no hard rules about what the similarities must be, or exactly *how* significant the overlap is (we will see below that this is an advantage of the theory).[95] If one or both refemes have different ascriptive content at different times, then overlap with respect to some times may be more important than others.

Recall the intuition that some refemes' reference depends upon, or is fixed by a relation with, other refemes' reference. I think it fairly intuitive that this relation of dependence cannot obtain if the two refemes are taken to be about radically different things (in the language I have developed, if they have very different ascriptive contents). My claim that reference inheritance requires ascription overlap is supposed to capture this intuition.

For example, suppose I hear you talking with great admiration about 'William'. Wrongly thinking I know who you mean, and holding a low opinion of the person I am thinking of, I say 'I didn't really like William when I met him!'. Since you and I are thinking of different people, surely we are talking of different people too. But your utterance of 'William' is in a causal chain with mine, and I would judge (as would anyone) that our utterances are of a kind. If this was all it took for reference to be inherited, then my utterance would inherit its reference from yours, and we would end up talking about the same person. Clearly this isn't happening because we think very different things about the two people we are referring to.

Let us see how this works in practice with Evans' Madagascar case. U3's and U2's ascriptions overlap significantly. The agent who produces U3 ascribes

[93]This is a simplification; see Burgess (2014) for a faithful discussion of the name's actual history. Nonetheless, this simplification has become a useful and standard test case for philosophy of language.

[94]Again, a historically false simplification, but let us stipulate for the sake of a good example.

[95]This is not a new idea. See Evans (1973; 1982 especially Chapters 5, 9 and 11) and Dickie (2011a).

*being an island*, *being off the south-east-coast of Africa* etc. The agent who produces U2 will ascribe very similar things. So U3 can inherit its reference from U2. U2 refers to the island, so U3 does as well. Neither U2's or U3's ascriptions overlaps significantly with U1's (because U1's utterer takes it to refer to an area of the mainland). So neither can inherit its reference from U1. U1 refers to the mainland, so there is no inconsistency.

Before moving on, let me discuss a potential objection to this part of the theory. Intuitively the sailors knowingly refer to the mainland (having correctly picked up the name from the locals), and Polo thinks he refers to an island. But if the sailors think they are referring to the mainland, and Polo thinks *he* refers to the island, then Polo's (refemes') ascriptions do not overlap significantly with the sailors' (refemes') ascriptions, and so Polo's refemes cannot inherit their reference from the sailors'. But it is generally thought that soon after their initial meeting, Polo co-refers with the sailors (at least for a while), even though he holds a false belief about the nature of the referent; we want to say that immediately after meeting the sailors, Polo is getting something *wrong*. But my theory seems to demonstrate that Polo refers to the island all along (or at least never co-refers with the sailors), and therefore that there is a sense in which Polo is *right*.

There are two problems with this objection. First, it assumes that the sailors' and Polo's ascriptions do not overlap significantly. But this isn't obvious.

On one hand it may simply be false that the sailors' and Polo's ascriptions do not overlap. I made clear above that my theory includes no reductive notion of 'significance' to determine when two refemes' ascriptions overlap significantly. Clearly the sailors' and Polo's ascriptions must overlap in *some* way (they both believe they are referring to a geographic location, often called 'Madagascar', etc.). It may well be that the best available notion of significant overlap applicable in this case would have the result that the sailors' ascriptions *do* overlap significantly with Polo's. For example, this would obtain if the most significant ascriptions the sailors or Polo make is the property of *being-called-'Madagascar'-by-the-Sailors* (compare Evans 1973 pp. 206-207 on deference). If the ascriptions do overlap, then the sailors' utterances will co-refer with Polo's (and thus both will initially refer to the mainland), though Polo wrongly thinks he refers to the island, and this part of the case will be as originally described.

On the other hand it may be indeterminate whether the sailors' ascriptions overlap significantly with Polo's. Suppose some but not all properties ascribed by the sailors are also ascribed by Polo, and vice versa. Whatever the relevant notion of significant overlap is, perhaps there is simply no fact of the matter as to whether it applies. Since it is indeterminate whether the ascriptions overlap (though I-relation and similarity both obtain) it will be indeterminate whether

Polo's utterances inherit from the sailor's, and indeterminate whether Polo refers to the island or the mainland. I discuss such cases further in the next section.

Now let us stipulate that the sailors' and Polo's ascriptions do not overlap significantly. What then? The original objection was that in this case, Polo's utterances could not inherit their reference from the sailors', and therefore Polo could not be wrong in the way he is generally thought to be: he would believe himself to be referring to an island, and he really *would* be referring to that island, whereas in the original case Polo is supposed to be mistaken (at least to be begin with).

The objection's second problem is that this isn't the only way to be mistaken. It is true that on my theory, Polo's belief that his utterances refer to the island is true. But he will still be mistaken in a different way. Polo believes that his utterances of 'Madagascar' co-refer with the sailors'. In the case as originally described, he was right about the co-reference, but wrong about referring to the island, because the sailors' utterances (with which he co-refers) refer to the mainland. In the case as my theory has it, and with the above stipulation, Polo will be right about referring to the island, but wrong about co-reference (since the sailors still refer to the mainland). Both readings have it that Polo gets something right, and something wrong. My reading isn't the standard, but I think it recovers the most important metasemantic intuitions.

#### 5.3.3.4 Indeterminacy

Reference and its inheritance are sometimes indeterminate. In this section I defend this general idea, enumerate three important ways this can happen, and show that my theory makes good predictions in each of the three kinds of case.

Let us begin with a simple motivational case, exhibiting several different kinds of indeterminacy. I attend a conference on a niche topic and meet two very similar, unremarkable men called 'John'. They both work on the same thing, hold similar views about it, and dress and speak in similar ways. I also meet a much more distinctive woman, Jackie. Our meeting is brief as she only stops by for the presentation given by one of the Johns.

Days after the conference, my memory runs the Johns together and I forget that there were different people. I token a cognitive refeme as if to think of 'the' man I met at the conference. Who, if anyone, does this refeme refer to? Certainly not to Jackie, Neil Armstrong, or anyone else aside from the Johns—this much is determinate. On one hand, it isn't determinate that it refers to either one of the Johns. On the other, it seems to have something to do with each of them, in a way it does not with Jackie, Neil Armstrong, or anyone else.

Later I run into Jackie. She talks to me about 'John', the man whose talk we both attended, and tells me that all those in the know call him 'Johnny'. In fact this is true of the man she knows (during whose presentation I met Jackie), but not of the other John I met. Believing myself to have only met one John, and that this is the man Jackie mentions, I pick up her use of 'Johnny' and join in the conversation, taking my utterances to co-refer with hers. Although I don't realise it, several of the things I ascribe to the putative referent of my utterances are false of the man Jackie is discussing (though they were true of the other John I met): that he gave a presentation in the morning, that he wore a red shirt, that he had a large moustache etc. Several others are true, and of these some were true of both Johns, and some of only the one Jackie knows. Given all this, it is clear that my utterances of 'Johnny' do not refer to anyone apart from the two Johns. But again, it is less clear who they *do* refer to, if anyone. On one hand, I have adopted Jackie's practice of using 'Johnny', and she clearly refers to one man in particular. My utterances are anaphoric on hers, so there is some evidence that I refer to him too. On the other, many of the things I believe about this man are false, and they are so because I have conflated him with someone else. Not only this but even some of my true beliefs are not 'rooted' only in him. All this considered, it is at least not obvious who I refer to, if anyone.

Finally, I write some notes in my diary on the conference. Years later, having forgotten the conference, the Johns and Jackie, I read my diary entry and reflect that John, whoever he was, must have been very boring. Given that my contemporary refeme inherits its reference, if any, directly from my diary entry, to whom does it refer?

Various parts of this case, in various ways, should motivate the idea that sometimes there just is no fact of the matter about what, if anything, a refeme refers to. In the rest of this section I'll make three claims about indeterminacy in reference inheritance, which I think make good sense of cases like the one I describe.

My first claim is that if all the conditions for a refeme A to inherit its reference from a refeme B are fulfilled, and also all the conditions for A to inherit from a refeme C, and B and C do not co-refer, then A's reference is indeterminate between the referents of B and C (or in other words: it is indeterminate whether A inherits its reference from B or from C). It *is* determinate that A does not refer to anything else besides B's or C's referent and also that A does not refer to both B's *and* C's referents, but it is indeterminate whether A refers to B's referent or to C's. I've described this with two refemes from which A might inherit, but a similar principle would apply for any number.

This analysis applies to the first part of the case above. During the conference I produced and observed many perceptual, cognitive and linguistic refemes referring to one of the Johns, and many referring to the other one. After I run together the two Johns in my mind, I token a single cognitive refeme. All the conditions are fulfilled for this refeme to inherit its reference from any of the refemes in the first group (all referring to one of the Johns) or any in the second group (all referring to the other), though not for it to refer to anything else. In particular: my eventual cognitive refeme is in a causal chain with many (perhaps all) refemes in both groups; I would judge it to be of a kind with many refemes in each of the two groups, and its ascriptive content will overlap significantly with that of most refemes in either group. By my first claim then, it is indeterminate whether I refer to one John or the other, though determinate that I do not refer to anyone else. This seems like the right analysis.

My second claim is that if two refemes share some but not all of their ascriptions, it may be indeterminate whether one inherits its reference from the other. My third condition on reference inheritance was that the ascriptive contents of two refemes must 'significantly overlap'. This term was deliberately vague. There are, I am sure, cases of definite overlap, but where there is no fact of the matter whether the overlap is *significant*.

This explains the second part of the case. My utterances of 'Johnny' are in a causal chain with Jackie's of the same name. And I would judge that my utterances are of a kind with hers. There is no question that Jackie's utterances determinately refer to just one man. But it is an open question whether *my* utterances refer to him. We now see why: the ascriptive content of my utterances is a borderline case of significant overlap with that of Jackie's, so they are a borderline case of inheritance. There is no simple fact of whether they inherit or not, and thus no simple fact of who they refer to.

My third claim is that if a refeme A definitely inherits its reference from a refeme B, and B's reference is indeterminate, then A's is indeterminate in the same way. If all the conditions are fulfilled for A to inherit its reference from B, then A and B must co-refer. If B's reference was indeterminate, then surely A's must be too. This is a very slight revision in the notion of co-reference; A and B do not co-refer in that they simply refer to the same thing, since B's reference is indeterminate. But of the available options, I think is the most straightforward reading of co-reference under indeterminacy.

This explains the last part of the case above. My refeme after reading the diary bears exactly the same indeterminacy as did my refemes at the time. But this is not because it has the features they did which made their reference indeterminate,

it is because it inherits its reference (and attendant indeterminacy) from them. This seems to me the correct result.

## 5.4 Solutions to the puzzles

In the last chapter I offered solutions to the puzzles which originally motivated this thesis. I said there that my solutions might strike some as a bit 'thin'—descriptions of the relevant kind of case which explain how they diverge from the standard 'good' cases (where referent and ascription match), but without explaining why these divergences arise in the first place. I promised that this would come in the next chapter. Now, with a theory of determination in hand, I can deliver on that promise. In this section I describe why and how various kinds of puzzling cases occur.

### 5.4.1 Alternative-satisfaction cases

These cases are the simplest. I claimed in Chapter 4 that an alternative satisfaction case is just a refeme with a certain referent and ascriptive content such that some object besides the referent satisfies the ascription (recall that these are independent of non-satisfaction cases, where the referent itself does not satisfy the ascriptions). It should be simple to see why these occur: neither primitive nor inherited reference requires or entails that a refeme's referent is the unique satisfier of that refeme's ascriptive content (inherited reference does not require that the referent satisfy the ascriptions at all). Since the properties ascribed by agents are often fairly minimal, and therefore not *uniquely* satisfied by anything, there will be many cases where the reference is fixed in spite of it not being the unique satisfier of the relevant ascriptive content. I'll give two concrete examples.

First, for primitive reference. You see a magpie in the distance, which you correctly identify as such. You token a visual refeme which refers to the magpie, and ascribe the property *being a magpie* (and since it is too far away to make out any details, you ascribe nothing else). You have identified the magpie by its magpie-ish features—at least those available to the naked eye—and so you ascribe only *being a magpie*. There are many other things which satisfy this ascription (many probably nearby); nevertheless you refer only to the one. Your refeme refers to that magpie, I claim, because that magpie caused your refeme to exist, because it does satisfy a significant portion of the properties you ascribe (in fact you only ascribe one, and it satisfies that), and because you are disposed to alter these ascriptions on the basis of changes in the magpie itself. All this is

fulfilled in virtue of only one object, so that is the referent, even though many others fulfil the ascriptive condition too.

Now for inherited reference. In the distant past, there was someone personally familiar with the great Roman orator Cicero. Utterances of the name 'Cicero' produced by this person will have been accompanied by very rich ascriptive content, perhaps so specific that only Cicero fulfilled it all. After Cicero's death, they continue using the name to discuss their late friend and his exploits. Each utterance they make is causally related to a previous one, with substantially overlapping ascriptions, and taken to be of a similar kind, so each refeme inherits its referential content (the man Cicero). Someone else hears one of these utterances, learns of Cicero, and goes on to speak of him to others. Having never met him, this second speaker knows much less about Cicero and his life, so their refemes have much less ascriptive content than the first speaker's. Even so, all the right relations hold for inheritance—in particular, the second speaker's ascriptions all overlap substantially with those of the first speaker. So each of the second speaker's refemes inherits its reference (in part) from those of the first speaker's (and in part from the earlier utterances by the second speaker). The cycle continues: a third, fourth, fifth, speaker and so on all learn of Cicero from the last one in the chain, and each time some ascriptions are 'lost'. Eventually this chain reaches me, a modern speaker of the name. I learn of Cicero only that he was a great Roman orator and when I token the name these are the only properties I ascribe. Even so, all the right relations hold for inheritance—in particular, my ascriptions all overlap substantially with those of the of a previous utterance (mine, or one of those by whoever taught me). Thus, I refer to Cicero, and only Cicero, even though all I ascribe is a property fulfilled by many different people.[96]

### 5.4.2 Non-satisfaction cases

Non-satisfaction cases are also fairly straightforward. These are refemes with a certain referent and ascriptive content such that the refeme does not satisfy (all of) the ascriptions. In most cases of this kind the referent satisfies *some*, but not all of the ascribed content. In particular, I argued above that if a refeme has its reference fixed primitively, a significant portion of its ascriptive content is satisfied by its referent. So the only non-satisfaction cases possible for refemes with primitive reference are those where some small portion of the ascriptive

[96]Of course, this story is substantially simplified. There would be more than one person involved at each stage, and some linguistic refemes would be written rater than spoken. Many of the speakers would probably seek out further information about Cicero and thus ascribe more than just what they inherit from a previous speaker. But these details are not to the point here.

content is not satisfied by the referent. These cases don't occur very often, but when they do it is because the 'link' between object and refeme-formation is not as strong as it could be. In inherited reference there is no such requirement, so non-satisfaction cases for refemes with inherited reference can include ascriptive content which is satisfied hardly or not at all. These occur much more often. Since the properties ascribed can vary substantially in long reference-inheriting chains of refemes (as in the case of Cicero) and since refemes with inherited reference need not ascribe *any* properties which their referents actually have, there are chains of refemes which preserve reference although some refemes ascribe little or nothing true of the referent. Again, I'll give two concrete examples in turn.

First for primitive reference. Perception is the standard case of primitive reference, and I have discussed several examples already. Suppose you are a keen bird watcher, usually able to distinguish willow tits from marsh tits, though the two look very similar. One day you see a willow tit close by, but by a trick of the light, its head feathers look darker and its neck thinner than they really are, and you mistake it for a marsh tit, usually distinguished by these features. You token a visual refeme and ascribe the properties *being a bird*, *being a tit*, *being a marsh tit* (and perhaps others). The visual refeme you token is caused by the bird, has ascriptive content mostly satisfied by the bird, and you are disposed to alter your ascriptions in line with alterations in the bird (if it flew away for instance, you would ascribe *flying*); your refeme refers to the bird. Nevertheless, its ascriptive content is not entirely satisfied; it is not a marsh tit, though you ascribed that property.

Most of the examples from elsewhere in the literature are cases of perceiving an object correctly, but in the wrong position within the visual field. (e.g. Shoemaker 1968; Campbell 2002) This position is taken to be a property, from which a non-satisfaction case can be described. I am wary of assuming that an object's apparent position is a *property*, like its apparent colour or opacity, so I have provided an example with colour and shape. This distinction shouldn't affect my overall argument though, and I will not argue for my wariness here.

Now for inherited reference. The most famous such non-satisfaction cases are Kripke's examples of Columbus, Gödel, and Peano. But these are all also alternative-satisfaction cases, which rather muddies the waters. Take instead the following plausible case, which I developed in §2.4. Alice attends a popular lecture on great problems in the history of mathematics. She hears about Fermat and his theorem, about Andrew Wiles who proved the theorem, and finally about the Collatz Conjecture which remains unproven. But things get muddled in Alice's memory—she leaves associating the name 'Andrew Wiles' with the property *prover of the Collatz Conjecture* and has forgotten everything else she learned

about him. If Alice utters 'Andrew Wiles' she will refer to the mathematician who *in fact* proved Fermat's last theorem, even though she doesn't ascribe this property, and *does* ascribe a property Wiles does not instantiate. Roughly, this is because Alice's refeme has a referential link going back to Wiles himself, and this is not severed by her false judgement about him.

### 5.4.3 Failure cases

Failure cases are refemes with no referent, but which behave in ways usually explained by a refeme's relation with its referent. I argued in Chapter 4 that this is not mysterious. When all goes as it should, a refeme's ascriptive content will be entirely made up of properties its referent actually has. Much of the refeme's 'behaviour' will be explained by this ascription. For instance, an informed speaker is much more likely to utter 'Andrew Wiles' in a conversation about mathematicians than one about artists, just because they ascribe *mathematician* but not *artist* when they token that name.

Why do failure cases occur? In this chapter I have given accounts of how reference and ascription are each determined. These mechanisms are such that a refeme can end up with some ascriptive content but no referent. These refemes will behave in a way which would be explicable (and would usually be explained) by their relation with a referent with these properties (as it should in the good cases, where referent and ascription cohere), even though they have no referent at all. These are the failure cases.

Perhaps the most famous example is 'Vulcan'. Le Verrier wrongly calculated that there must exist a planet between Mercury and the Sun with certain mass and orbital characteristics, and tokened a cognitive refeme which he took to refer to it. His refeme had ascriptive content *being a planet*, *being positioned between Mercury and the Sun* etc.. In 'arriving' at this cognition Le Verrier resembles the detective of §5.3.2.2. The refeme has its ascription fixed primitively: Le Verrier infers that something must have these certain characteristics, so when he tokens the refeme he A-judges that its referent has each of these characteristics. This process neither entails nor requires that the refeme has a referent. Given these ascriptions, whether it does have a referent is also fixed primitively. If it did, Le Verrier's refeme would refer to something which caused the refeme to exist, fulfilled a significant portion of the ascriptions, and with respect to which he was disposed to update these ascriptions. There is no such thing, so the refeme doesn't refer (contrast with the detective, whose procedure is identical, but who does end up referring).

This explains the 'behaviour' of the refeme. If the refeme *did* refer, probably[97] it would refer to a planet, so locutions like 'Vulcan is a planet' are sensible from Le Verrier's point of view (even if actually false). Le Verrier would be much more likely to token 'Vulcan' in a discussion about astronomy than one about artists, because of the ascriptions he makes when tokening that name.

### 5.4.4 Frege's puzzle

Millar characterises Frege's puzzle as a mistake of one object for two objects, because the one is presented in two different ways. Or as I claimed in Chapter 4, Frege's puzzle is any case involving two different refemes, with the same referent but different ascriptions, in which an agent does not realise that the two refemes co-refer and therefore makes differing ascriptions (or it is the contrast between this and a case without such a mistake). With the terminology I have developed in this chapter, we could say that the agent *should* make a CR-judgement that the two refemes co-refer, and on that basis should update their ascriptions in step, but in fact does not. These cases occur because the same referent can be determined for two refemes with entirely different ascriptions. I describe three cases below.

First, for primitive reference. Recall my adaption of Wettstein's case, where you are backstage at a rock concert, hoping to catch the star. As you stand in the dressing room, you seem to see two different people facing in opposite directions. In fact you see the one rock star, reflected in their own mirrors. You token one visual refeme for one reflection, and ascribe the properties it presents. Let us say that the rock star's right hand side is done up in bright colours, glitter and huge fake lashes. This refeme will ascribe the properties *brightly made up* and *wearing fake lashes* etc.. Their left side is very plain and stark. Your refeme for the other reflection ascribes the properties *poorly made up*, *pale*, etc.. Both refemes actually refer to the star, because both refemes were caused by them, both refemes' ascriptive content is entirely satisfied by them, and in both cases the agent is disposed to alter their ascriptions following alterations in the star's actual appearance. Nevertheless, because the ascriptions are different, the agent doesn't judge the two refemes to co-refer (or equivalently: doesn't judge that one person *is* the other).

This case is straightforward partly because you never realise that the 'two' stars are one. But Frege cases can also emerge with respect to an object which one

[97]Recall that primitive reference only requires a significant portion of the ascriptions to be satisfied by an object. In a case where Le Verrier was right, except that the massive object was not a planet (but a comet or other solar object), his tokens of 'Vulcan' would still refer, but not to a planet. Such cases are very unlikely though.

initially identifies between two refemes, and it may be instructive to consider one.

I work at a carnival, in the mirror maze. Unlike the punters, I am well used to the odd visual effects of the different mirrors and can usually navigate the maze with ease. One day, I am walking through the maze alone carrying a torch. Used to the mirrors, I am able without difficulty to keep track of the tactile refeme referring to the torch in my hand, and the various visual refemes for its various reflections, and maintain running CR-judgements of the co-reference of all of these. Suddenly I am momentarily disoriented and think there may be someone else around carrying a similar torch. Though I don't lose track of any of the torch's reflections, the 'link' between them and the torch in my hand is broken. Focusing on a certain reflection I ask myself 'is *this* torch [in my hand] the same as *that* one [in the mirror]?'. This is a case of Frege's puzzle: the two refemes really do co-refer, but I don't behave as if they do (at least not after I've been spooked). The analysis is the same here as before: I do not judge that the two refemes co-refer, and so do not ascribe the same things to each. Before I was spooked, I would have updated the ascriptive content of the visual refeme (for the reflection) on the basis of observed changes in the torch I felt in my hand. If the torch started to vibrate to signal a low battery, then I would have added *low battery* to the ascriptive content of both the tactile refeme (for the torch in my hand) and the visual refeme (for the reflection), precisely on the basis that they co-refer. Once I am spooked, I won't co-ascribe in this way. If the torch only vibrates after I have wondered, then I will only add *low battery* to the tactile refeme. Moreover, I will make different ascriptions to the 'two' torches. In order to form the thought when I wonder if they are the same, I must somehow distinguish one torch from the other: I will add something like *seen in the mirror* to the ascriptive content of the visual refeme and withhold it from the tactile one, and I will withhold the (previously ascribed) *held in my hand*.[98] The need for square-bracketed adjectives in the question I ask myself above is a good illustration of this phenomenon.[99]

Now for inherited reference. By a process similar to that in the rock star case, ancient astronomers thought they had seen two separate stars on separate occasions—one in the morning, and one in the evening—when in fact they had merely seen Venus twice. They dubbed each 'star' with a separate name. When

[98]Or, if I was previously operating so automatically that I was not thinking of the fact that I held the torch in my hand, I will now add *held in my hand* to the ascriptive content of the tactile refeme, but will withhold it from the visual refeme. The effect is the same—that the ascriptive contents differ.

[99]This is one case in the which a structural feature of the sentence used to notate a thought is a good guide to the features of the thought itself.

uttering either name, the astronomers would ascribe several properties that Venus has (*being a celestial body*, *being bright*, *appearing in the morning sky* or *appearing in the evening sky* respectively, etc.) and at least one it does not (*being a star*). Utterance of these names would thus constitute partial non-satisfaction cases, similar to Alice's utterances of 'Wiles'; in spite of the few bad ascriptions, the utterances still both refer to Venus. But because the ascriptions differ, the astronomers would not judge that their utterances co-refer.

#### 5.4.4.1 'Paderewski' cases

Since 'Paderewski' cases are just an instance of Frege's puzzle, this analysis applies to them as well, though the explanation is slightly more involved. Peter initially learns the name as identifying a famous pianist. He will make utterances of 'Paderewski' ascribing *being a pianist* and not ascribing *being a statesman*. These utterances all refer to the actual Paderewski in the standard way: they are all in the same I-relation, which reaches back before Peter's uses, they all have overlapping ascriptions, and Peter would take them all to be of a like kind. Later, Peter learns the name again, this time as identifying a famous statesman. Believing that pianists are not usually statesmen, he will make utterances of 'Paderewski'. These too are all I-related (some may even be I-related to the other set of utterances), they all have overlapping ascriptions, and Peter would take them all to be of a like kind. Crucially though, the ascriptions of the utterances in the first set differ significantly from those in the second set, so there is no inheritance of reference between the two practices. In spite of this, all the utterances refer to the same man. In the case I considered earlier, Peter reminds himself (mistakenly) of the non-identify of 'one' Paderewski with the 'other' by saying 'Paderewski is not Paderewski'. We now see how the content of each utterance of the name is determined. First, Peter makes an utterance of the name, I-related to those in the first set, and ascribing *being a pianist*. He takes it to co-refer with earlier utterances in that set, and since he believes (correctly) that such utterances ascribed *being a pianist*, he ascribes that again here. He does not ascribe *being a statesman*. The utterance's ascriptions thus overlap significantly with earlier ones of the first kind, and it inherits from them. Then Peter says 'is not', and then utters 'Paderewski' again. Now the situation is reversed: he takes this utterance to co-refer with earlier utterances of the second kind, and as such ascribes *being a statesman*, but not *being a pianist*. Unknown to him, this ends up determining the very same referent as his preceding utterance of the same name.

### 5.4.5 Transformation

Transformation cases are a converse to Frege's puzzle. In Frege's puzzle, we mistake one object for two; when the magician 'changes' a king into an ace, we mistake two objects for one. Why?

Our perceptual system uses properties to keep track of objects; as a similarly-shaped and coloured complex (or 'gestalt') moves through space, the object file we have assigned to it 'sticks' to it. This tracking is robust under some changes in the 'object' tracked (like shape or colour), and even under brief occlusion by other objects, so long as the original gestalt remains.[100]

In the trick, the magician switches one card for another so fast that the overall gestalt *does* remain in vision, though some features change—namely those indicating the card's value. Since the gestalt is constant, the same object file tracks the 'one' card, but since the features have changed, the ascriptive content attached to this file is updated. Before the switch, the ascriptive content is satisfied to the extent that the object file refers to the card. This is also true after the switch, though in virtue of different content and properties. All the other conditions for primitive reference hold too, so both before and after the switch, the object file refers to whichever card is visible. Like the other puzzles, this is case of things not going as they 'should'.

## 5.5 Conclusion

This concludes my discussion of how the content of particular refemes is determined. For reference and ascription separately, I have argued for a split model, where some refemes have their content fixed 'primitively', and some by their relations with other refemes. We have seen that the mechanisms determining ascription and reference are closely related, though not in the way many have previously assumed.

[100]This process need not involve *ascription* of these properties to an object, or the manipulation of such ascriptions in a tracking process. It may do, or it may be an entirely subpersonal mechanism in which elements of the visual system just react automatically to certain changes of stimulus. See *inter alia* Pylyshyn and Storm (1988), Pylyshyn (2001), Feldman and Tremoulet (2006), and Campbell (2012).

# Chapter 6

# Conclusion

Let us see where we have come. In the introduction I distinguished three kinds of significance things might have: external content, internal content, and content *simpliciter*. The debate at the heart of much analytic philosophy has been over whether the first or the second is identical with the third. The claim of Unity was that whatever the answer is, it is the same for language, perception and thought.

In Chapter 2 I argued that this claim is true. I showed that the structure the historical debate on this topic has actually taken is identical with that which exhausts what can be said about the Unity claim, and that this fact was itself a good reason to believe Unity. As I did there, I stress that this is a defeasible reason. My project is not significant because I take on all comers, but because it is the first to engage with this question directly.

If there is a unified kind of content, there must be a unified kind of thing which has it. In Chapter 3 I introduced refemes and gave my theory of their nature; refemes of diverse kinds turned out not to be so diverse after all.

In Chapter 4 I answered the central question and gave a theory of the nature of content. I argued that both external and internal content play a role—that content is a complex of reference and ascription. Unlike previous theorists, I gave a detailed theory of what ascription *is*, and rejected the idea that reference must be determined by internal content.

This left a gap in the theory, to be filled by the last chapter. I argued for twin, related views of how reference and ascription are determined, both inspired by the primitive/inherited structure of traditional casual theories of reference. This part of my account is the most powerful. As I hope I showed, it explains in detail the exact dynamics of individual refemes and their content, and how these are embedded in huge, long lived networks of communication and content change.

To return to the sentiment of the introduction: my theory is highly conjectural, and probably open to all sorts of counterexamples. Nonetheless, it is supported by the available evidence, and is specific and detailed enough to be usefully held up to scrutiny.

# References


Adams, Fred and Robert Stecker (1994). "Vacuous Singular Terms". In: *Mind and Language* 9.4, pp. 387–401. DOI: 10.1111/j.1468-0017.1994.tb00314.x.

Almog, Joseph (2005). "Is a Unified Description of Language-and-Thought Possible?" In: *Journal of Philosophy* 102.10, pp. 493–531. DOI: 10.5840/jphil20051021030.

Almog, Joseph, John Perry, and Howard Wettstein, eds. (1989). *Themes from Kaplan*. Oxford University Press, pp. 565–614.

Altham, J. E. J. (1973). "The Causal Theory of Names. II". In: *Aristotelian Society Supplementary Volume* 47.1, pp. 187–225. DOI: 10.1093/aristoteliansupp/47.1.187.

Arstila, Valtteri and Kalle Pihlainen (2009). "The Causal Theory of Perception Revisited". In: *Erkenntnis* 70.3, pp. 397–417. DOI: 10.1007/s10670-008-9153-7.

Assadian, Bahram and Giorgio Sbardolini (2023). "Performative reference". In: *Synthese* 202.2. DOI: 10.1007/s11229-023-04284-2.

Austin, J. L. (1965). *How to do things with words*. Clarendon Press.

Bach, Kent (2010). "Getting a Thing into a Thought". In: *New Essays on Singular Thought*. Ed. by Robin Jeshion. 1. Oxford: Oxford University Press. Chap. 1, pp. 39–63. DOI: 10.1093/acprof:oso/9780199567881.003.0002.

Batty, Clare (2010). "Olfactory Experience II: Objects and Properties". In: *Philosophy Compass* 5.12, pp. 1147–1156. DOI: 10.1111/j.1747-9991.2010.00352.x.

Båve, Arvid (2008). "A deflationary theory of reference". In: *Synthese* 169.1, pp. 51–73. DOI: 10.1007/s11229-008-9336-4.

Beck, Jacob (2014). "Analogue Magnitudes, the Generality Constraint, and Nonconceptual Thought". In: *Mind* 123.492, pp. 1155–1165. DOI: 10.1093/mind/fzu154.

Bertolet, Rod (1980). "Demonstratives and intentions". In: *Philosophical Studies* 38.1, pp. 75–78. DOI: 10.1007/bf00354528.

— (1986). "Referring, Demonstrating, and Intending". In: *Philosophy Research Archives* 12, pp. 251–260. DOI: 10.5840/pra1986/1987121.

Brandom, Robert (1998). *Making It Explicit. Reasoning, Representing, and Discursive Commitment*. Harvard University Press.

Braun, David (1993). "Empty Names". In: *Noûs* 27.4, p. 449. DOI: 10.2307/2215787.

— (2005). "Empty Names, Fictional Names, Mythical Names". In: *Noûs* 39.4, pp. 596–631. DOI: 10.1111/j.0029-4624.2005.00541.x.

Burch, Robert (1979). "James and the 'New' Theory of Reference". In: *Transactions of the Charles S. Peirce Society* 15.4.

Burge, Tyler (1974). "Truth and Singular Terms". In: *Noûs* 8.4, p. 309. DOI: 10.2307/2214437.

— (1979). "Sinning Against Frege". In: *The Philosophical Review* 88.3, p. 398. DOI: 10.2307/2184957.

— (2009). "Perceptual Objectivity". In: *The Philosophical Review* 118.3, pp. 285–324. DOI: 10.1215/00318108-2009-001.

Burgess, J. P. (2014). "Madagascar revisited". In: *Analysis* 74.2, pp. 195–201. DOI: 10.1093/analys/anu036.

Campbell, John (2002). *Reference and Consciousness*. Oxford University Press. DOI: 10.1093/0199243816.001.0001.

— (2012). "Susanna Siegel's the Contents of Visual Experience". In: *Philosophical Studies* 163.3, pp. 819–826. DOI: 10.1007/s11098-012-0013-6.

Chastain, Charles (1975). "Reference and Context". In: *Language, Mind, and Knowledge*. Ed. by Keith Gunderson. Minnesota Studies in the Philosophy of Science.

Chisholm, Roderick M. (1977). "Thought and its Reference". In: *American Philosophical Quarterly* 14.2, pp. 167–172.

Crane, Tim (2001). "Intentional Objects". In: *Ratio* 14.4, pp. 336–349. DOI: 10.1111/1467-9329.00168.

— (2005). "What is the Problem of Perception?" In: *Synthesis philosophica* 20.2, pp. 237–264.

— (2006). "Is There a Perceptual Relation?" In: *Perceptual Experience*. Ed. by Tamar Szabo Gendler and John Hawthorne. Oxford University Press, pp. 126–146. DOI: 10.1093/acprof:oso/9780199289769.003.0004.

— (2009). "Intentionalism". In: *The Oxford Handbook of Philosophy of Mind*. Ed. by Ansgar Beckermann, Brian P. McLaughlin, and Sven Walter. Oxford University Press, pp. 474–493. DOI: 10.1093/oxfordhb/9780199262618.003.0029.

— (2011a). "Existence and quantification reconsidered". In: *Contemporary Aristotelian Metaphysics*. Ed. by Tuomas E. Tahko. Cambridge University Press, pp. 44–65. DOI: 10.1017/cbo9780511732256.005.

Crane, Tim (2011b). "Singular Thought. I". In: *Aristotelian Society Supplementary Volume* 85.1, pp. 21–43. DOI: 10.1111/j.1467-8349.2011.00194.x.

— (2012). "What is the Problem of Non-Existence?" In: *Philosophia* 40.3, pp. 417–434. DOI: 10.1007/s11406-011-9354-1.

— (2013). *The Objects of Thought*. Oxford: Oxford University Press.

Crane, Tim and Craig French (2021). "The Problem of Perception". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward Zalta.

Cullison, Andrew and Ben Caplan (2010). "Descriptivism, scope, and apparently empty names". In: *Philosophical Studies* 156.2, pp. 283–288. DOI: 10.1007/s11098-010-9589-x.

deRosset, Louis (2020). "Reference Switch and Bad Dubbings". URL: https://www.uvm.edu/~lderosse/madagascar.pdf (visited on 02/2024).

Devitt, Michael (1981). *Designation*. Columbia University Press. DOI: 10.7312/devi90836.

Dickie, Imogen (2011a). "How Proper Names Refer. III". In: *Proceedings of the Aristotelian Society* 111.1pt1, pp. 43–78. DOI: 10.1111/j.1467-9264.2011.00298.x.

— (2011b). "Visual Attention Fixes Demonstrative Reference By Eliminating Referential Luck". In: *Attention: Philosophical and Psychological Essays*. Ed. by Christopher Mole, Declan Smithies, and Wayne Wu. Oxford University Press.

Donnellan, Keith S. (1966). "Reference and Definite Descriptions". In: *The Philosophical Review* 75.3, p. 281. DOI: 10.2307/2183143.

Dummett, Michael (1973). *Frege. Philosophy of language*. Duckworth.

— (1993). *Origins of Analytical Philosophy*. Duckworth.

Echeverri, Santiago (2015). "Object Files, Properties, and Perceptual Content". In: *Review of Philosophy and Psychology* 7.2, pp. 283–307. DOI: 10.1007/s13164-015-0275-y.

Edelberg, Walter (1986). "A new puzzle about intentional identity". In: *Journal of Philosophical Logic* 15.1. DOI: 10.1007/bf00250547.

Evans, Gareth (1973). "The Causal Theory of Names. I". In: *Aristotelian Society Supplementary Volume* 47.1, pp. 187–225. DOI: 10.1093/aristoteliansupp/47.1.187.

— (1982). *The Varieties of Reference*. Ed. by John McDowell. Oxford: Oxford University Press.

— (1985). "Things Without the Mind". In: *Collected Papers*. Ed. by Antonia Phillips. Oxford: Clarendon Press. Chap. 9, pp. 249–291. DOI: 10.1093/oso/9780198236276.001.0001.

Everett, Anthony (2003). "Empty Names and 'Gappy' Propositions". In: *Philosophical Studies* 116.1, pp. 1–36. DOI: 10.1023/b:phil.0000005533.25543.36.

Feldman, Jacob and P. Tremoulet (2006). "Individuation of visual objects over time". In: *Cognition* 99.2, pp. 131–165. DOI: 10.1016/j.cognition.2004.12.008.

Fetzer, J. H. and P. Humphreys, eds. (1998). *The New Theory of Reference: Kripke, Marcus, and its origins*.

Field, Hartry (1973). "Theory Change and The Indeterminacy of Reference". In: *The Journal of Philosophy* 70.14, p. 462. DOI: 10.2307/2025110.

— (1994). "Deflationist Views of Meaning and Content". In: *Mind* 103.411, pp. 249–285. DOI: 10.1093/mind/103.411.249.

Fodor, J. A. (1984). "Semantics, Wisconsin style". In: *Synthese* 59.3, pp. 231–250. DOI: 10.1007/bf00869335.

Foster, John (2004). "The Problem Of Perception". In: *The Harvard Review of Philosophy* 12.1, pp. 4–18. DOI: 10.5840/harvardreview20041215.

Frege, Gottlob (1948). "On Sense and Reference". In: *The Philosophical Review* 57.3, pp. 209–230. DOI: 10.2307/2181485.

— (1956). "The Thought: A Logical Inquiry". In: *Mind* 65.259, pp. 289–311.

— (1980). *Philosophical and Mathematical Correspondence*. Ed. by Gottfired Gabriel et al. Trans. from the German by Hans Kaal. Blackwell.

Gauker, Christopher (2015). "The Illusion of Semantic Reference". In: *On Reference*. Ed. by Andrea Bianchi. Oxford University Press, pp. 21–39. DOI: 10.1093/acprof:oso/9780198714088.003.0002.

Geach, Peter (1967). "Intentional Identity". In: *The Journal of Philosophy* 64.20, p. 627. DOI: 10.2307/2024459.

— (1969). "The Perils of Pauline". In: *The Review of Metaphysics* 23.2.

— (1972). *Logic Matters*. Oxford: Blackwell.

Gentry, George (1943). "Reference and Relation". In: *The Journal of Philosophy* 40.10, p. 253. DOI: 10.2307/2017371.

Gómez-Torrente, Mario (2019). *Roads to Reference*.

Goodman, Rachel (2016). "On the supposed connection between proper names and singular thought". In: *Synthese* 195.1, pp. 197–223. DOI: 10.1007/s11229-016-1202-1.

Goodman, Rachel and James Genone, eds. (2020). *Singular Thought and Mental Files*. Oxford University Press. DOI: 10.1093/oso/9780198746881.001.0001.

Gray, Aidan (2015). "Lexical Individuation and Predicativism about Names". In: *Thought: A Journal of Philosophy* 4.2, pp. 113–123. DOI: 10.1002/tht3.164.

Green, E. J. (2018). "What Do Object Files Pick Out?" In: *Philosophy of Science* 85.2, pp. 177–200. DOI: 10.1086/696376.

Green, E. J. and Jake Quilty-Dunn (2021). "What Is an Object File?" In: *The British Journal for the Philosophy of Science* 72.3, pp. 665–699. DOI: 10.1093/bjps/axx055.

Grice, H. P. (1957). "Meaning". In: *The Philosophical Review* 66.3, p. 377. DOI: 10.2307/2182440.

— (1961). "The Causal Theory of Perception". In: *Aristotelian Society Supplementary Volume* 35.1, pp. 121–168. DOI: 10.1093/aristoteliansupp/35.1.121.

— (1969). "Vacuous Names". In: *Words and Objections. Essays on the Work of W.V. Quine*. Ed. by Donald Davidson and K. J. J. Hintikka. Springer. DOI: 10.1007/978-94-010-1709-1.

— (1991). "Presupposition and Conversational Implicature". In: *Studies in the Way of Words*. Harvard University Press.

Hanks, Peter (2015). *Propositional Content*. Oxford University Press. DOI: 10.1093/acprof:oso/9780199684892.001.0001.

Hawthorne, John and David Manley (2012). *The Reference Book*. Oxford University Press.

Heagren, Hugo (2026). "On baptisms". In: *Inquiry*. DOI: 10.1080/0020174X.2026.2618510.

Horwich, Paul (1988). *Meaning*. Oxford University Press.

Hughes, Justin (1984). "Group speech acts". In: *Linguistics and Philosophy* 7.4, pp. 379–395. DOI: 10.1007/bf00631073.

Inwagen, Peter van (1997). "Creatures of Fiction". In: *American Philosophical Quarterly* 14.4, pp. 299–308.

Jackman, Henry (2020). "Temporal externalism, conceptual continuity, meaning, and use". In: *Inquiry* 63.9–10, pp. 959–973. DOI: 10.1080/0020174x.2020.1805706.

Jeshion, Robin (2000). "Ways of Taking a Meter". In: *Philosophical Studies* 99.3, pp. 297–318. DOI: 10.1023/a:1018712718908.

— (2010a). "Introduction to New Essays on Singular Thought". In: *New Essays on Singular Thought*. Ed. by Robin Jeshion. Oxford: Oxford University Press, pp. 1–36. DOI: 10.1093/acprof:oso/9780199567881.003.0001.

— ed. (2010b). *New Essays on Singular Thought*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199567881.001.0001.

Johnston, Mark (2004). "The Obscure Object of Hallucination". In: *Philosophical Studies* 120.1–3, pp. 113–183. DOI: 10.1023/b:phil.0000033753.64202.21.

Kahneman, Daniel and Anne Treisman (1984). "Changing views of attention and automaticity". In: *Varieties of Attention*. Ed. by D. R. Davies and R. Parasuraman. Academic Press series in cognition and perception. Academic Press.

Kahneman, Daniel, Anne Treisman, and Brian J Gibbs (1992). "The reviewing of object files: Object-specific integration of information". In: *Cognitive Psychology* 24.2, pp. 175–219. DOI: 10.1016/0010-0285(92)90007-o.

Kaplan, David (1968). "Quantifying in". In: *Synthese* 19.1–2, pp. 178–214. DOI: 10.1007/bf00568057.

— (1989a). "Afterthoughts". In: *Themes from Kaplan*. Ed. by Joseph Almog, John Perry, and Howard Wettstein. Oxford University Press, pp. 565–614.

— (1989b). "Demonstratives. An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals". In: *Themes from Kaplan*. Ed. by Joseph Almog, John Perry, and Howard Wettstein. Oxford University Press, pp. 565–614.

— (1990). "Words". In: *Aristotelian Society Supplementary Volume* 64.1, pp. 93–120. DOI: 10.1093/aristoteliansupp/64.1.93.

Katz, Jerrold J. (1994). "Names Without Bearers". In: *The Philosophical Review* 103.1, p. 1. DOI: 10.2307/2185871.

— (2001). "The End of Millianism: Multiple Bearers, Improper Names, and Compositional Meaning". In: *Journal of Philosophy* 98.3, pp. 137–166. DOI: 10.5840/jphil20019833.

Kim, Jaegwon (1977). "Perception and Reference without Causality". In: *The Journal of Philosophy* 74.10, p. 606. DOI: 10.2307/2025913.

King, Jeffrey C. (2024). "Structured Propositions". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward Zalta.

Koch, Steffen and Alex Wiegmann (2022). "Folk Intuitions about Reference Change and the Causal Theory of Reference". In: *Ergo an Open Access Journal of Philosophy* 8.0. DOI: 10.3998/ergo.2226.

Kripke, Saul (1979). "A Puzzle about Belief". In: *Meaning and Use*. Ed. by Avishai Margalit. Springer Netherlands, pp. 239–283. DOI: 10.1007/978-1-4020-4104-4_20.

— (1980). *Naming and Necessity. Lectures Given to the Princeton University Philosophy Colloquium*. Cambridge, MA: Harvard University Press.

Lackey, Jennifer (2017). "Group Assertion". In: *Erkenntnis* 83.1, pp. 21–42. DOI: 10.1007/s10670-016-9870-2.

Lewis, David (1984). "Putnam's paradox". In: *Australasian Journal of Philosophy* 62.3, pp. 221–236. DOI: 10.1080/00048408412340013.

Linnebo, Øystein (2012). "Reference by Abstraction". In: *Proceedings of the Aristotelian Society* 112 (1), pp. 45–71. DOI: 10.1111/j.1467-9264.2012.00325.x.

Locke, John (1975). *An Essay Concerning Human Understanding*. Ed. by Peter Hidditch. Oxford: Clarendon Press.

Ludwig, Kirk (2013). "Proxy Agency in Collective Action". In: *Noûs* 48.1, pp. 75–105. DOI: 10.1111/nous.12013.

— (2020). "What are group speech acts?" In: *Language & Communication* 70, pp. 46–58. DOI: 10.1016/j.langcom.2019.04.004.

Martin, M. G. F. (2002a). "Particular Thoughts & Singular Thought". In: *Royal Institute of Philosophy Supplement* 51, pp. 173–214. DOI: 10.1017/s1358246100008134.

— (2002b). "The Transparency of Experience". In: *Mind and Language* 17.4, pp. 376–425. DOI: 10.1111/1468-0017.00205.

McDowell, John (1977). "On the Sense and Reference of a Proper Name". In: *Mind* LXXXVI.342, pp. 159–185. DOI: 10.1093/mind/lxxxvi.342.159.

— (1982). "Truth-Value Gaps". In: *Logic, Methodology and Philosophy of Science VI. Proceedings of the Sixth International Congress of Logic, Methodology, and Philosophy of Science*. Ed. by L. J. Cohen. Amsterdam: North Holland Publishing Co., pp. 299–133.

McGinn, Colin (1981). "The mechanism of reference". In: *Synthese* 49.2, pp. 157–186. DOI: 10.1007/bf01064297.

McGrath, Matthew and Devin Frank (2024). "Propositions". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward Zalta.

Michaelson, Eliot (2023a). "The Vagaries of Reference". In: *Ergo* 9. DOI: 10.3998/ergo.3115.

— (2023b). "Unspeakable names". In: *Synthese* 201.2. DOI: 10.1007/s11229-023-04038-0.

Michaelson, Eliot and Marga Reimer (2022). "Reference". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward Zalta.

Millar, Boyd (2015). "Frege's Puzzle for Perception". In: *Philosophy and Phenomenological Research* 93.2, pp. 368–392. DOI: 10.1111/phpr.12230.

Nowak, Ethan and Eliot Michaelson (2020). "Who's Your Ideal Listener?" In: *Australasian Journal of Philosophy* 99.2, pp. 257–270. DOI: 10.1080/00048402.2020.1741658.

Nudds, Matthew (2009). "What Are Auditory Objects?" In: *Review of Philosophy and Psychology* 1.1, pp. 105–122. DOI: 10.1007/s13164-009-0003-6.

O'Callaghan, Casey (2008). "Object Perception: Vision and Audition". In: *Philosophy Compass* 3.4, pp. 803–829. DOI: 10.1111/j.1747-9991.2008.00145.x.

Ogden, C. K. and I. A. Richards (1923). *The Meaning of Meaning*.

Orilia, Francesco and Michele Paolini Paoletti (2025). "Properties". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward Zalta.

Pears, D. F. (1976). "The causal conditions of perception". In: *Synthese* 33.2–4, pp. 25–40. DOI: 10.1007/bf00484708.

Pepp, Jessica (2012). "Reference and Referring: a framework". In: *Reference and Referring*. Ed. by William P. Kabasenche, Michael O'Rourke, and Slater Matthew H. The MIT Press, pp. 1–32. DOI: 10.7551/mitpress/9581.003.0002.

Perry, John (1977). "Frege on Demonstratives". In: *The Philosophical Review* 86.4, p. 474. DOI: 10.2307/2184564.

— (2001). *Reference and Reflexivity*. Stanford: CLSI Publications.

Price, Henry Habberly (1961). *Perception*. London: Methuen & Co.

Prosser, Simon (2019). "The Metaphysics of Mental Files". In: *Philosophy and Phenomenological Research* 100.3, pp. 657–676. DOI: 10.1111/phpr.12577.

Putnam, Hilary (1973). "Meaning and Reference". In: *Journal of Philosophy*. DOI: 10.2307/2025079.

Pylyshyn, Zenon W. (2001). "Visual indexes, preconceptual objects, and situated vision". In: *Cognition* 80.1–2, pp. 127–158. DOI: 10.1016/s0010-0277(00)00156-6.

— (2007). *Things and Places: How the Mind Connects with the World*. MIT Press. DOI: 10.7551/mitpress/7475.001.0001.

Pylyshyn, Zenon W. and Ron W. Storm (1988). "Tracking multiple independent targets: Evidence for a parallel tracking mechanism". In: *Spatial Vision* 3.3, pp. 179–197. DOI: 10.1163/156856888x00122.

Quine, W. V. O. (1948). "On What There Is". In: *The Review of Metaphysics* 2.5, pp. 21–38.

— (1956). "Quantifiers and Propositional Attitudes". In: *The Journal of Philosophy* 53.5, p. 177. DOI: 10.2307/2022451.

— (1960). *Word and Object*. Massachusetts: MIT Press.

Recanati, François (2012). *Mental Files*. Oxford University Press. DOI: 10.1093/acprof:oso/9780199659982.001.0001.

Reimer, Marga (1991). "Do demonstrations have semantic significance?" In: *Analysis* 51.4, pp. 177–183. DOI: 10.1093/analys/51.4.177.

— (1992). "Three views of demonstrative reference". In: *Synthese* 93.3, pp. 373–402. DOI: 10.1007/bf01089275.

Russell, Bertrand (1905). "On Denoting". In: *Mind*.

— (1919). *Introduction to Mathematical Philosophy*. George Allen & Unwin.

— ed. (1956a). *Logic and Knowledge. Essays 1901-1950*. George Allen & Unwin.

— (1956b). "The Philosophy of Logical Atomism. Essays 1901-1950". In: ed. by Robert C. Marsh. George Allen & Unwin.

— (1970). "Knowledge by Acquaintance and Knowledge by Description". In: *The Problems of Philosophy*. Oxford University Press. Chap. 5.

Sainsbury, Mark, ed. (2003a). *Departing from Frege*. Routledge. DOI: 10.4324/9780203218938.

— (2003b). "Fregean Sense". In: *Departing from Frege*. Ed. by Mark Sainsbury. Routledge, pp. 125–136. DOI: 10.4324/9780203218938.

— (2003c). "Names, fictional names, and "really"". In: *Departing from Frege*. Ed. by Mark Sainsbury. Routledge. DOI: 10.4324/9780203218938.

— (2003d). "Sense without Reference". In: *Departing from Frege*. Ed. by Mark Sainsbury. Routledge. DOI: 10.4324/9780203218938.

— (2005). *Reference Without Referents*. Oxford: Oxford University Press.

— (2010). "Intentionality without Exotica". In: *New Essays on Singular Thought*. Ed. by Robin Jeshion. Oxford: Oxford University Press. Chap. 10, pp. 300–318. DOI: 10.1093/acprof:oso/9780199567881.003.0011.

— (2014). "The Same Name". In: *Erkenntnis* 80.S2, pp. 195–214. DOI: 10.1007/s10670-014-9705-y.

Sainsbury, Mark and Michael Tye (2012). *Seven Puzzles of Thought and How to Solve Them. An Originalist Theory of Concepts*. Oxford University Press. DOI: 10.1093/acprof:oso/9780199695317.001.0001.

Salmon, Nathan (1986). *Frege's Puzzle*.

— (1998). "Nonexistence". In: *Noûs* 32.3, pp. 277–319. DOI: 10.1111/0029-4624.00101.

Schellenberg, Susanna (2010). "The particularity and phenomenology of perceptual experience". In: *Philosophical Studies* 149.1, pp. 19–48. DOI: 10.1007/s11098-010-9540-1.

— (2016). "Perceptual Particularity". In: *Philosophy and Phenomenological Research* 93.1, pp. 25–54. DOI: 10.1111/phpr.12278.

Scholl, Brian J. and Zenon W. Pylyshyn (1999). "Tracking Multiple Items Through Occlusion: Clues to Visual Objecthood". In: *Cognitive Psychology* 38.2, pp. 259–290. DOI: 10.1006/cogp.1998.0698.

Scholl, Brian J., Zenon W. Pylyshyn, and Jacob Feldman (2001). "What is a visual object? Evidence from target merging in multiple object tracking". In: *Cognition* 80.1–2, pp. 159–177. DOI: 10.1016/s0010-0277(00)00157-8.

Searle, John R. (1958). "Proper Names". In: *Mind* LXVII.266, pp. 166–173. DOI: 10.1093/mind/lxvii.266.166.

— (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press. DOI: 10.1017/cbo9781139173438.

— (1983). *Intentionality*. Cambridge University Press. DOI: 10.1017/CBO9781139173452.

Segal, Gabriel (1989). "The Return of the Individual". In: *Mind* XCVIII.389, pp. 39–57. DOI: 10.1093/mind/xcviii.389.39.

Shahmoradi, Ayoob (2024). "Referring with or without Individuating the Referent". In: *The Philosophical Quarterly*.

Shoemaker, Sydney S. (1968). "Self-Reference and Self-Awareness". In: *The Journal of Philosophy* 65.19, pp. 555–567. DOI: 10.2307/2024121.

Siegel, Gabriel (2022). "Perceptual Modes of Presentation as Object Files". In: *Erkenntnis*. DOI: 10.1007/s10670-022-00633-8.

Siegel, Susanna (2005). "The Contents of Perception". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward Zalta.

Smith, Barry and Thomas Bittner (2001). "A Unified Theory of Granularity, Vagueness, and Approximation". In: *COSIT Workshop on Spatial Vagueness, Uncertainty and Granularity*.

Snowdon, Paul (1981). "Perception, Vision and Causation". In: *Proceedings of the Aristotelian Society* 81.1, pp. 175–192. DOI: 10.1093/aristotelian/81.1.175.

— (1992). "How to interpret 'direct perception'". In: ed. by Tim Crane. Cambridge Univesity Press. Chap. 3, pp. 48–78. DOI: 10.1017/cbo9780511554582.004.

Snowdon, Paul and Howard Robinson (1990). "The Objects of Perceptual Experience". In: *Aristotelian Society Supplementary Volume* 64.1, pp. 121–166. DOI: 10.1093/aristoteliansupp/64.1.121.

Soames, Scott (2002). *Beyond Rigidity. The Unfinished Semantic Agenda of Naming and Necessity*. Oxford University Press. DOI: 10.1093/0195145283.001.0001.

Soteriou, Matthew (2000). "The Particularity of Visual Perception". In: *European Journal of Philosophy* 8.2, pp. 173–189. DOI: 10.1111/1468-0378.00107.

Stampe, Dennis W. (1977). "Toward a Causal Theory of Linguistic Representation". In: *Midwest Studies in Philosophy* 2, pp. 42–63. DOI: 10.1111/j.1475-4975.1977.tb00027.x.

Strawson, P. F. (1964). *Individuals*. Methuen.

Sullivan, Arthur (2010). "Millian Externalism". In: *New Essays on Singular Thought*. Ed. by Robin Jeshion. Oxford: Oxford University Press, pp. 246–270. DOI: 10.1093/acprof:oso/9780199567881.003.0009.

Taylor, Isaac (1896). *Names and Their Histories. Alphabetically Arranged as a Handbook of Historical Geography and Topographical Nomenclature*. Percival Rivington.

Taylor, Kenneth A. (2010). "On Singularity". In: *New Essays on Singular Thought*. Ed. by Robin Jeshion. Oxford: Oxford University Press. Chap. 3, pp. 77–102. DOI: 10.1093/acprof:oso/9780199567881.003.0004.

Thau, Michael (2002). *Consciousness and Cognition*. New York: Oxford University Press. DOI: 10.1093/oso/9780195141818.001.0001.

Thomasson, A. L. (2003). "Fictional Characters and Literary Practices". In: *The British Journal of Aesthetics* 43.2, pp. 138–157. DOI: 10.1093/bjaesthetics/43.2.138.

Tiedke, Heidi (2011). "Proper Names and their Fictional Uses". In: *Australasian Journal of Philosophy* 89.4, pp. 707–726. DOI: 10.1080/00048402.2010.514930.

Tye, Michael (2007). "Intentionalism and the Argument from no Common Content". In: *Philosophical Perspectives* 21.1, pp. 589–613. DOI: 10.1111/j.1520-8583.2007.00137.x.

Vendler, Zeno (1976). "Thinking of Individuals". In: *Noûs* 10.1, p. 35. DOI: 10.2307/2214473.

Wettstein, Howard (1984). "How to bridge the gap between meaning and reference". In: *Synthese* 58.1, pp. 63–84. DOI: 10.1007/bf00485362.

— (1986). "Has Semantics Rested on a Mistake?" In: *The Journal of Philosophy* 83.4, p. 185. DOI: 10.2307/2026531.

White, Alan R. (1961). "The Causal Theory of Perception". In: *Aristotelian Society Supplementary Volume* 35.1, pp. 121–168. DOI: 10.1093/aristoteliansupp/35.1.121.

Wittgenstein, Ludwig (2009). *Philosophical Investigations*. German and English. Ed. by P. M S. Hacker and Joachim Schulte. Trans. by G. E. M. Anscombe, P. M S. Hacker, and Joachim Schulte. 4th ed. Wiley and Blackwell.

Zalta, Edward (1988). *Intensional Logic and Metaphysics of Intentionality*. MIT Press.

— (2001). "Fregean Senses, Modes of Presentation, and Concepts". In: *Philosophical Perspectives* 15, pp. 335–359.

Zarepour, Mohammad Saleh (2018). "On Crane's Psychologistic Account of Intentionality". In: *Acta Analytica* 33.4, pp. 453–462. DOI: 10.1007/s12136-018-0342-y.

Ziff, Paul (1977). "About Proper Names". In: *Mind* LXXXVI.343, pp. 319–332. DOI: 10.1093/mind/lxxxvi.343.319.