This dissertation is submitted for the degree of Doctor of Philosophy

[This is a copy of my doctoral thesis at Cambridge. The supervisor was Simon Blackburn and the degree was awarded in 2014.]

The Strange Case of the Missing

Theory of Reference

JP Smit

Darwin College

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

The length of the dissertation, including footnotes and bibliography, does not exceed the permitted length.

The Strange Case of the Missing Theory of Reference – JP Smit

In the thesis I present a novel theory of semantic reference, which I call the *coordination view* of semantic reference.

In chapter one I develop a novel theory of conventions, which I call the *coordinating rule view* of conventions. On the coordinating rule view, roughly, a convention is a rule which originates in a meta-coordination game and is adopted in order to deal with a series of future coordination games. This view is Lewisian in spirit, but rejects Lewis's view that a convention is a regularity in behaviour.

In chapter two I use the coordinating rule view of conventions in order to define the coordination view of semantic reference. On the coordination view, roughly, a name N semantically refers to a particular o if, and only if, there is a convention to use N to speaker-refer to o. I also argue that a consideration of the coordination view of semantic reference makes it plain that Kripke's causal theory has no non-trivial explanatory content. The causal chain that exists between a baptism and downstream use of a name comes about in virtue of the fact that knowledge of the content of a convention is typically causally acquired. This fact, however, is a mere triviality about conventions as such, not a revelation about reference. I also argue that non-Kripkean versions of causalism have little to recommend them.

In chapter three I compare the coordination view of semantic reference with the traditional descriptivism of Russell and Frege. I defend the claim that traditional descriptivism is plainly false. This follows from the commitment of the traditional descriptivists to what I call *eccentricity* about names, i.e. the view that the descriptive condition which *the utterer* of the name associates with the name determines the semantic reference of the name. In fact, I claim that traditional descriptivism is so plainly false that we need to explain why Russell and Frege proposed it. I try to provide such an explanation.

In chapter four I address various issues raised by the discussion in the previous three chapters. The main claim defended is that it is very strange that nothing like the coordination view of semantic reference features in discussion concerning semantic reference, given that the analogous view of sentence meaning, i.e. the view that sentence meaning is to be defined in terms of speaker-meaning and convention, has received a lot of attention.

Table of contents

Introduction

This thesis concerns the semantic reference of names in natural language. Discussion on this topic has generally taken the form of a debate between descriptivism and causalism. Descriptivism is historically most closely associated with the work of Russell and Frege, who I will call the 'traditional descriptivists', whereas causalism is most closely associated with the work of Kripke. On the standard view of these debates these two positions are rival attempts to answer a single question, namely 'how does a name refer to a thing?'.

I will propose a novel theory of the semantic reference of names and present a new view of the work of Kripke and the traditional descriptivists. I will claim that Kripke's causal theory has no non-trivial content and that traditional descriptivism is so blatantly false that we need to explain how such fine philosophers defended such a strange doctrine. I will argue that these claims can be demonstrated by considering the nature of conventions.

I will reserve the term 'semantic reference' for the topic Kripke addressed in *Naming and Necessity*. It is a basic fact about the kind of thing Kripke wrote about, i.e. semantic reference, that it is determined by convention. In chapter one I develop an alternative to the Lewisian view of conventions. This theory I call the coordinating rule view of conventions. The coordinating rule view stays true to Lewis' basic insight that conventions and game theoretical coordination games are conceptually closely related, but replaces Lewis' regularities and preferences with rules and goals. On the coordinating rule view, a convention is best understood as a rule which originates in a meta-coordination game about future coordination games.

In chapter two I use the coordinating rule view of conventions, coupled with a view on which communication consists in the transmission of information, to define semantic reference in terms of convention and speaker's reference. I call this the coordination view of semantic reference. On the coordination view, a name semantically refers to a specific particular if, and only if, all members of a linguistic community are disposed to, absent defeaters and false beliefs, employ a rule that advises using the name to speaker-refer to the specific particular. I also try and show, partly by comparing Kripkean causalism to the coordination view and partly by looking at considerations concerning conventions in general, that Kripke's causal theory has no non-trivial content. My main complaint is that, on any theory of convention, people have to learn the content of a convention from other people. Hence we should expect users of a convention to typically stand in some causal relationship to the event whereby the convention was introduced. This follows purely in virtue of the epistemic role of such chains. Any conceivable theory of semantic reference will have the consequence that such causal-his-

torical chains exist. The existence of such chains is trivial and shows nothing about reference; we also have such chains linking us back to the people who established the convention of driving on the left-hand side in the UK.

Traditional descriptivism, of course, does not claim that such causal-historical chains exist. In chapter three I argue that traditional descriptivism is so plainly false that we need an explanation of how such fine philosophers proposed such a strange doctrine. My main argument depends on the fact that Russell and Frege both committed to the idea that the semantic reference of a name is determined by speakers individually. Yet it is obvious that the use of names is a matter of interpersonal convention, i.e. is subject to interpersonal coordination. I argue that the only way to make sense of traditional descriptivism is to interpret Russell and Frege as confusing the semantic reference of names with the *beliefs* people have about the semantic reference of names.

In chapter four I address various topics that arose naturally from the preceding discussion but are best left till when my case has been fully made. The main issue discussed concerns the relation of the coordination view of semantic reference to similar, and influential, theories of sentence meaning. I also explore some different ways of thinking about semantics.

The main theme running though this thesis is that the method that semanticists have generally used to in order to adjudicate between the views of Kripke and those of the traditional descriptivists is defective. Semanticists have mostly followed the so-called 'method of cases', whereby we are supposed to explain intuitions concerning the truth-conditions of utterances. Clearly the method of cases should be a part of our tool box; it would be impossible to determine the semantics of a natural language without it. The problem, however, is that insufficient attention has been paid to what these intuitions are intuitions *about*. This is something that we can only learn from theoretical reflection about the kinds of entities relevant to communication.

Theoretical reflection on the kinds of entities relevant to communication can be done by considering transcendental question like what must be the case in any possible communicative interaction. This method, of course, has its dangers. Our imagination is limited and fallible; we often learn that something is possible only by discovering that it is actual. Not using such methods, however, also has its dangers. If we don't focus on such abstract questions we can become blinded by the assumptions that we do have and may systematically misinterpret data. Below I will briefly run through an explanation of the kind of theoretical reflection I am talk-

ing about. This will also serve to give the reader a flavour of the basic view underlying this thesis.

In any theory we have to start somewhere, start by assuming that the notion of the speaker's reference of a name is theoretically sound. In other words, start by assuming that speakers try to bring particulars to the attention of other speakers by performing communicative acts. Any speaker who wishes to bring some particular to the attention of the hearer has to provide the hearer with some publicly available cues as to the particular he wishes to bring to the hearer's attention. The hearer must then use such cues as evidence as to what the speaker has in mind. Define the 'public referent' of a communicative act as the object which best fits all the publicly available evidence, whether intentionally provided or not, as to the speaker's referent[1].

Here we can note two things. First, we can construe such a communicative interaction as a coordination game where any state in which the speaker's referent and the object that the hearer believes to be the speaker's referent coincide is an equilibrium, i.e. a state where no agent has an incentive to unilaterally change his behaviour. Second, the fact that the public referent is to be determined in terms of external evidence means that the speaker's referent and the public referent can diverge. In other words, the speaker can be wrong about the thing that best fits the relevant evidence. Which thing best fits such cues is an empirical fact and no-one has infallible access to empirical facts.

If two people reach an equilibrium in one coordination game then this equilibrium is salient to future coordination games and can become a convention between the two people. In chapter one I argue that conventions are best thought of as rules that subjects adopt in order to deal with coordination games that they think will recur. The main concern when choosing such a rule, of course, is whether others will follow the same rule, and so the choice of such a rule amount to a coordination game about future coordination games, i.e. a meta-coordination game. The notion of a conventional referent can be thought of as a referent determined in virtue of a rule which originates in such a meta-coordination game. This meta-coordination game takes place under extreme informational constraints as the future is unknown. As the future unfolds one may have to break these rules in specific cases, without thereby necessarily abandoning these rules as a general policy.

---

[1] The notion of a public referent is theoretically useful. In Smit (2012) I argue that bare demonstratives are best viewed as terms that, on an occasion of use, have a speaker's referent and a public referent, but no conventional referent.

If the above construal of communication is roughly correct, then it follows that the content of conventions are empirical facts in the same sense as which rules people follow are empirical facts. Speakers need to have beliefs about the content of these rules and such beliefs can be wrong. In order to fully explain a communicative interaction we now need to distinguish between the speaker's referent, the conventional referent and the public referent. We also need to keep track of the parties' beliefs about what these entities are. It is only once we have done so that we can usefully apply the method of cases, as the method of cases cannot be applied without having some sense of what would count as evidence that a subject misunderstood a question about 'what someone said by uttering a sentence *s*' as a question about speaker's reference, public reference, or some other *species* of meaning that is not conventional reference. In fact, if the views defended in this thesis are correct then problem does not only apply to experiments where non-philosophers are asked questions about what utterances meant. Rather it is simply impossible that anyone, even a brilliant semanticist, could ever have consistently Russellian intuitions about semantic reference. A philosopher who claims to have such intuitions must have been misled into thinking about the wrong sort of thing in virtue of thinking in terms of some ill-defined, amorphous entity like 'what is said' or 'what is expressed' by an utterance. Once we think of actual cases with a proper understanding of the needed distinctions the identity of the semantic referent in such cases will tend to be obvious. In chapter four I argue that, even in cases where it is not clear what we should take the semantic referent of a name to be, as in exaggerated 'Gödel'/'Schmidt' cases, the difficulty cannot be understood as a clash between Russellian and Kripkean intuitions. Rather what is at stake in such cases is the exact nature of the equilibrium selection rule governing our semantic behaviour.

If there is an enemy in this thesis it is notions like 'what is said' or 'what is expressed'. Or, alternatively, the notion 'semantic reference' where this is not used as a simple synonym for 'conventionally determined reference', but as something derivable from judgments about 'what is said' or 'what is expressed'. I have no idea what these notions are supposed mean. This is not to say that I don't have intuitions about 'what is said' by an utterance, just that I have no idea what these intuitions are supposed to be about. In this thesis I defend the idea that semantics has paid a heavy price for using notions like these.

Chapter 1: An alternative to the Lewisian view of conventions

1. Introduction

In *Convention* (1969), Lewis set out to investigate the platitude that language is conventional. His key idea is that conventions are regularities that result from recurrent coordination games where the parties to the coordination game are aware of the status of the resultant regularity. In this chapter I develop an alternative view. I will claim that conventions are not regularities, but rules. A convention exists in a society when the members of a society are disposed to follow such rules. I will also present an alternative view of how conventions are sustained and defend the idea that conventions should be modelled as arising in virtue of a meta-coordination game, i.e. a coordination game about a series of future coordination games.

My analysis has two parts. In the first part I will defend my proposal by working through three objections to Lewis's definition of convention that motivate departing from his views in the ways that I propose. In the second part I will propose various amendments to Lewis' underlying theory of conventions. Note that, while the view developed here will depart from Lewis in various ways, it is still very much in the spirit of Lewis' own analysis. I believe that Lewis' insight that conventions are conceptually linked to what may be termed coordination games is the single most important fact about conventions.

2. Lewis's definition of convention and the nature of the present inquiry

Lewis defines the notion of a convention as follows:

> A regularity R in the behaviour of members of a population P when they are agents in a recurrent situation S is a *convention* if and only if it is true that, and it is common knowledge in P that, in any instance of S among members of P,
> (1) everyone conforms to R;
> (2) everyone expects everyone else to conform to R;
> (3) everyone has approximately the same preferences regarding all possible combinations of actions;
> (4) everyone prefers that everyone conform to R, on condition that at least all but one conform to R;

(5) everyone would prefer that everyone conform to *R′*, on condition that at least all but one conform to *R′*,

where *R′* is some possible regularity in the behaviour of members of *P* in *S*, such that no one in any instance of *S* among members of *P* could conform both to *R′* and to *R* (Lewis, 1969: 76).

Lewis later on (1969: 78) develops a definition that allows for exceptions to the strict conditions given above. I will, as most of those writing about Lewis do, focus mainly on the exceptionless version. The core of Lewis' theory is that conventions are regularities that arise in response to game-theoretical coordination problems where participants are aware of the status of the resultant regularity. The defining characteristic of a 'game' as such is that the context of interaction must be *strategic*, i.e. optimal strategy for one party must depend on the behaviour of the other party (or parties) involved. A game is a game of coordination if the interests of the actors are *aligned*, i.e. actors have the same ordinal ranking of the different outcomes[2]. Lewis does not, however, require that the parties to a convention have perfectly aligned preference. Rather, in condition (3), he merely requires that preferences must be 'approximately' the same (1969: 76). This allows him to treat games like 'battle of the sexes'[3] as potentially giving rise to conventions (1969: 14), even though the preferences of the parties involved give rise to differing ordinal rankings of the possible outcomes.

The basic constraint on the game theoretical structure of the interaction that Lewis does impose depends on his ingenious notion of a *coordination* equilibrium. The standard notion of an equilibrium in game theory is that of a set of strategies such that no actor can make himself better off by unilaterally changing his strategy. Such strategies are said to be in equilibrium as, if the actors somehow hit on such a combination, the outcome thereby reached is likely to be stable. Lewis's notion of a coordination equilibrium differs from that of a standard equilibrium in that that, in a coordination equilibrium, no actor can be made better off by *any* actor unilaterally changing their strategy. Mutual defection in a prisoner's dilemma[4], for instance, is

---

[2] Lewis follows Schelling (1960) in distinguishing between games of pure coordination and games of pure conflict (1969: 13-14).

[3] In a typical 'battle of the sexes', two parties have to choose between going out and staying in. Both would prefer the outcomes where they choose the same option to those where they choose different outcomes. One party, however, would prefer both parties going out to both parties staying in, and the other party would prefer both parties staying in to both going out.

[4] In a typical two-person prisoner's dilemma parties have to choose between 'defecting' and 'cooperating'. Both would prefer to be the sole defector, whereas both need to avoid being the sole co-operator. It is further stipulated that both prefer mutual cooperation to mutual defection. The sole equilibrium in a one-shot prisoner's dilemma is mutual defection.

an equilibrium, but not a coordination equilibrium, as one party can make the other better off by cooperating. Using this notion of a coordination equilibrium, Lewis ultimately imposes the requirement that the kind of coordination game needed to give rise to a convention must be a game with multiple coordination equilibria (1969: 16).

One way of understanding the conceptual question as to the nature of a convention is to understand it as an analysis of everything that we apply the natural language term 'convention' to. The basic goal of such an enquiry would be to try and arrive at an analysis that gives necessary and sufficient conditions that apply to everything that a competent speaker of English would call a convention. This, however, is not what I will be trying to do. It is far from clear that the things that we call a 'convention' form a kind. The things that we most commonly think of when we think of conventions include the matter of driving on the same side of the road in a given country and linguistic rules like referring to Kripke as 'Kripke'. My analysis will apply well to such cases, but there are also other forms of behaviour that we call conventions that it does not straightforwardly apply to. Many will claim that, 'in some sense', fashion is a matter of convention, that working until five o' clock is a matter of convention, that proper table manners and other matters of etiquette are conventional, and so on. While, for instance, how we use language and how we choose to dress may be related in some non-trivial way, it seems *prima facie* unlikely that we are here dealing with exactly the same kind of behaviour. Lewis, similarly, was not trying to give an account that would capture every single use of the word 'convention'. Where his theory clashed with ordinary usage he was untroubled, admitting the existence of "genuine usages that do not fall under my analysis"[5] (Lewis, 1976: 113). My goal here is similar. I will not be trying to give a theory everything that can be called a 'convention', i.e. trying to do the job of a lexicographer.

Lewis wrote *Convention* with the aim of arriving at a theory of conventions that can be of use in the study of language. The theory he arrives at aims primarily at, and is tailor-made for, capturing what we may call 'the conventionality of language'. I will proceed similarly. I will attempt to develop a notion of convention that is important, with this importance being a matter of being useful in explaining our use of language. Hence, in the first instance, the task is not one of analysing what we *mean* by 'conventions', but developing a notion that captures at least some of what makes an action one that accords with what we call a 'convention', and is

---

[5] Lewis did suggest that such usage may be derivative in some way, i.e. to depend on his notion of convention in some deep way (Lewis, 1976: 113). I suspect the same of the view developed here.

useful in explaining linguistic behaviour. This task can be glossed as being a matter of 'explaining the sense in which language is conventional'. For this reason I will continue to talk of giving a theory of 'convention', where this task is understood as explained here. I take it that, as explained above, this is also what Lewis was trying to do.

The difference between what I will try to do and an analysis that captures how we use the term 'convention' should not be overstated. Where possible I will try to develop a theory that is consistent with how we use the term 'convention'. The only difference is that, if our usage differs from the theory to be developed here, but the way in which usage differs has little or no explanatory value when it comes to language or introduces needless complication, I will ignore common usage. Where such considerations do not apply, however, I will try to make the theory consistent with common usage. In fact, I think that the view to be defended here sticks closer to how we apply the term 'convention' than Lewis's view does.

3. Conventions as coordinating rules

3.1 Coordinating rules

The task to be carried out then, is this: develop a theory of 'convention' that explains our linguistic behaviour and departs from common usage of the term 'convention' only when necessary. In order to state my view in an intuitive form, several notions need to be defined, the first being the notion of a 'rule'.

> Rule: A rule R is an injunction to perform an action K if specified circumstances C obtain. Rules can be stated in the form 'if C obtains, perform K'.

Rules are employed by agents in order to promote various goals. In this way the rules of valid inference promote truth-preservation, the codified rules of food labelling promotes consumer safety, some self-imposed rules of conduct promote productivity, and so on. For current purposes, a goal of action must be distinguished from a mere benefit that is realised in virtue of action. Take for, instance, someone who takes up jogging in order to improve his health. Further stipulate that the person enjoys meeting the kind of people one meets through jogging, but this consequence of jogging would not have been sufficient to motivate him to take up jogging. Ordinarily, we would not object if the person said that he jogs in order to improve his

health and meet people. On the definition used here, however, only the health benefits of jogging count as a goal of action, the latter is a mere benefit.

We can allow for cases of self-deception or a lack of self-knowledge by not requiring that the goal the person thinks motivates his behaviour actually be the goal that motivates his behaviour, or that the person be aware of the goal that motivates his behaviour. Consider someone who drives on the left-hand side of the road, thinks that he does so due to religious conviction, and does not realize that, if driving on the left did not enable him to avoid head-on collisions, he would switch to driving on the right. Such a person essentially has an incorrect theory about why he persists in a specific course of action. In such a case the goal of obeying a divine injunction does not, for our purposes, counts as a goal of action, while the goal of avoiding head-on collisions does. On the final analysis then, we require that the goal[6] of an action must be a potential outcome of the action which explains why the action is committed, independently of whether the person is aware of the fact that the outcome motivates his action.

All rules promote a goal via a certain *mechanism*, i.e. in a certain way. For example, the way in which the rule 'if at the office, don't use the internet' promotes productivity is that eliminates one source of distraction, the way in which the rule 'if you experience severe pain while exercising, stop' promotes being healthy is that it stops those who follow it from exacerbating a serious injury, and so on. One mechanism whereby a rule can promote a goal is coordination. Intuitively, a rule that promotes a goal via coordination does so in virtue of making it come about that our actions are similar or differ in some relevant way. More precisely, we can define the notion of a *coordinating rule* as follows:

> A rule *R* is a coordinating rule if, and only if,
>> (1) *R* is followed in order to promote a goal *G;*
>> (2) the effectiveness in promoting *G* of an action that exhibits *R* in a strategic context of interaction *C* primarily depends on the number of actions in *C* that exhibit *R*;
>> (3) the effectiveness in promoting *G* of an action that exhibits *R* in *C* increases, all else being equal, as a function of the number of actions in *C* that exhibit *R*;

---

[6] For an alternative view that defines conventions in terms of (entities like) rules and goals, as opposed to regularities and preferences, see Miller (1992). Miller defines conventions in terms of 'procedures' and 'collective ends' (1992: 436 – 437). Though the substance of my theory differs a lot from his, my choice of terminology is not supposed to be indicative of any deep difference on the nature or role of goals and rules.

(4) if all actions in *C* exhibit *R*, then there is no action in *C* such that, if replaced by an action that violates *R*, the replacement action would have been more effective in promoting *G*; and

(5) if all actions in *C* exhibit *R*, then there is no action in *C* such that it would have been more effective in promoting *G* if some combination of the other actions in *C* violated *R*.

For present purposes, define the notion of behaviour 'exhibiting a rule' as behaviour that accords with what the rule prescribes, independently of whether we would class the behaviour as rule-following or not. Define the notion of a strategic context as a situation in which the optimal action to perform depends on what other actions will be performed. Note that condition (4) is an adaptation of the standard idea of an equilibrium, whereas condition (5) is an adaptation of Lewis' idea of a coordination equilibrium. Including condition (4) allows us to exclude rules that advise us to cooperate in prisoner's dilemmas. This is required as such rules are not conventions, but moral norms. Including condition (5) allows us to exclude rules that advise us to defect in prisoner's dilemmas. This is required as, at least in one-off cases, defection is optimal independently of how others behave, whereas, in the case of conventions, the best way to promote a relevant goal is conditional on how others behave.

Note that the above definition is not stated in terms of agents who perform actions, but instead in terms of actions themselves. This is done as, strictly speaking, conventions only require a multiplicity of actions that are strategically related, not a multiplicity of agents. Hence we should include cases where the interactive context is an intertemporal one where different actions of the same individual promote some goal in virtue of all such actions being actions that exhibit the same rule. Consider a being who must eat once a day and who can minimize his chances of falling ill by spacing these meals as far apart as possible. This implies that the being should eat at the same time every day, but it does not matter when he does so. If such a being adopts a rule 'every day, eat at noon', then this counts as a coordinating rule, even if only one person[7] is involved. Note that, by the same standard, a secret script that an individual invents in order to keep his diary entries private also counts as a set of coordinating rules.

---

[7] Alternatively, we could have defined conventions as holding between 'agents', where agents are objects that can be persons or time-slices of persons.

Most relevant contexts of interaction will feature different agents with each performing an action, and, as formulating such cases in terms of the actions themselves can be a bit inelegant, I will mostly formulate my claims in terms of agents performing actions. So, turning to some everyday examples, the rule 'if in the UK, drive on the left' is a coordinating rule on the above definition as, first, people do so in order to avoid head-on collisions; second, the effectiveness of driving on the left in avoiding head-on collisions primarily depends on how many people also drive on the left when I encounter them; third, the effectiveness of driving on the left in avoiding head-on collisions increases, all else being equal, as a function of how many people also drive on the left when I encounter them; fourth, no person can improve his chances of avoiding a head-on collision by driving on the right; and, fifth, no person, or group of persons, can improve the chances of avoiding a head-on collision for anyone who drives on the left, by driving on the right. Equivalent claims are true for the way in which linguistic conventions promote communication, the way in which adopting a given currency lowers transaction costs, the way in which having the initial caller call back when a call is dropped aids speedy resumption of the call, and so on.

Note that the above construal of conventions as coordinating rules is not incompatible with the truism that conventions often advise us to all do different things. Even in such cases we all still 'do the same thing' in the sense of acting in a way that *exhibits the same rule*. Conventions that advise us to all do different things can be phrased as rules that all can follow, so that it is natural to say we all do the same thing. Even when we say that everyone in the UK 'drive on the same side', this only makes sense if we implicitly interpret 'same side' in terms of egocentric coordinates like left and right. The whole point of such a convention is to ensure that vehicles going in opposite directions drive on *different* sides of the road. If everyone really only ever used the same side, where 'same' is defined without reference to egocentric coordinates, it would result in catastrophe.

3.2 Defeasibility of coordinating rules

The view I defend is that conventions are coordinating rules. If this is accurate, then the conditions under which a convention can be said to exist are the conditions under which we can say that someone follows a coordinating rule. To do this we need to first pay attention to the conditions under which someone who should count as a rule-follower of the required type would have a good reason *not* to follow the rule. First we need to distinguish between two kinds of defeasibility, call the first 'internal defeasibility'.

> A rule R is internally defeasible if, and only if, R is followed in order to promote a goal G and there can be occasions where violating R is more effective than following R in promoting G.

All rules are not, of course, internally defeasible. The rules of valid inference cannot be over-ruled in the required manner by some feature of a specific context. A coordinating rule, however, must always be internally defeasible. Consider the matter of driving on the left-hand side of the road in the UK. As people generally obey it, it is rational to adopt the rule 'if driving in the UK, stick to the left-hand side'. There can, however, be occurrences of driving in the UK in which this is no longer an optimal way of not-crashing. I can always encounter some other driver who, either by mistake or not, drives on the right and effectively forces me to break the rule. This is true of all coordinating rules. The efficacy of coordinating rules depends on other people obeying them as well. This means that there can always be a scenario where others' breaking the rule creates a situation where I have a *prima facie* reason to break the rule. Hence all coordinating rules are internally defeasible rules.

Coordinating rules are also defeasible in a broader sense, which I will call 'external defeasibility'.

> A rule R is externally defeasible if, and only if, it is followed in order to promote a goal G and there can be occasions where some goal G' is more motivating than G and achieving G' necessitates violating R.

Consider cases where the specific circumstances forces my hand in some way by, for instance, providing a non-strategic reason for action. If, for example, I am driving in the UK and there is no car within a mile from me, but there is a giant pothole in front of me, I have a reason to drive on the right in order to get around the pothole. In such a case the goal of not-crashing does not guide my behaviour, as the goal of avoiding the pothole is more motivating. The motivating power of a coordinating rule is always externally defeasible as there can always arise a context in which some other goal is more motivating. Hence coordinating rules are always internally and externally defeasible.

3.3 Convention defined

With the above notions defined and explained, a relatively simple statement of the conditions under which a convention exists can be given.

> A rule R is a convention among a sub-group S of a population P, if, and only if, R is a coordinating rule that, absent external or internal defeaters, and absent relevant false beliefs, all members of S are disposed to follow.

Call the above view the *coordinating rule view of conventions*. Some of the reasons for adopting it should already be clear from the above discussion. The clause about the absence of relevant false beliefs is included in order to deal with cases where people try to coordinate their behaviour, but happen to be mistaken in some relevant way. In other words, cases where some person is disposed to drive on the right in the UK in virtue of thinking that driving on the right is the current, dominant driving standard in the UK, or a case where someone uses 'Lucas' to speaker-refer to Krugman in virtue of thinking that it is standard usage, etc. In such case, even though the behaviour of the person will not generate the regularity that allows him to be party to a 'convention', as Lewis defines it, there is a sense in which the person is party to the convention as he is still trying to coordinate his behaviour with others, despite being unsuccessful in virtue of his mistake concerning the dominant standard.

Below I will discuss the objections to Lewis' account that motivate rejecting his view in favour of the coordinating rule view.

4. Objections to Lewis' definition of 'convention'

4.1 Objection 1: The knowledge requirement.

Lewis, in his characterization of conventions, requires that his conditions (1) – (5) must be common knowledge among the parties to the convention. This requirement is convincingly criticized in Burge (1975)[8]. Burge points out that we can imagine speakers who are only aware of the existence of one language and believe that the words in the language are somehow 'naturally', or by supernatural fiat, connected to what they mean (1975: 250). Such speakers would not understand their own language use as conditional on how others use lan-

---

[8] Burge's argument has generally been found to be persuasive. See, for example, the discussion in Blackburn (1984: 120 -122).

guage. Yet we would not hesitate to call their language use conventional, despite the fact that they will explicitly deny its conventional nature.

Burge also points out a deeper problem (1975: 250-251). Throughout the history of philosophy, many have claimed that certain values and beliefs, thought to be somehow natural, are actually based, in some deep sense, on conventions. In this way certain basic doctrines in mathematics, logic and ontology have been claimed to reflect human conventions, as opposed to how things objectively are. On Lewis' construal, such a move would seem to be inherently absurd, as it would be constitutive of conventions that those who use them understand their continued use to be conditional on other people also conforming to the same convention. The claim such an argumentative move is intrinsically incoherent is implausible.

The coordinating rule view does not require the parties to a convention to understand the state that they find themselves in. As such it does not, by definitional *fiat*, rule out the possibility discovering that some aspect of our behaviour is conventional. Note that this is not to deny that common knowledge has a fundamental role in to play in explaining the origin or persistence of conventions, as clearly it does. All that is denied is that the parties to a convention need to understand why they act as they do.

4.2 Objection 2: Rules instead of regularities

Below I will argue that conventions are not regularities, but rules. Before we get to the meat of the argument, note one initial point in favour of such a claim. Conventions, as a quick google search will conform, are commonly said to be the kinds of things we can follow or violate. If conventions are rules of a certain kind, then this matter of common usage is explained as rules are also commonly said to be the kinds of things we can follow or violate. Regularities, however, are not the kinds of things that are commonly said to be followed or violated. The expressions 'follow a regularity' and 'violate a regularity' are simply not standard English. We can, of course, speak of 'actions in accord with a regularity', as we can speak of 'actions in accord with a convention'. But, we can equally well speak of 'actions in accord with a rule', and hence this latter usage does not favour the regularity-view over the rule-view. The basic point is that we typically portray conventions as things that can be followed or violated. This matter of usage is explained by rule-view of conventions, but not on a regularity-view of conventions. This fact should serve to give the rule-view some initial plausibility.

The main aim in this chapter, however, is not to explain matters of usage, but to develop a notion of convention that is of use in explaining phenomena like language. To this end, consider the regularity that is supposed to be constitutive of the existence of a convention. In the case of conventions concerning driving, this is a matter of the side of the road that different people choose to drive on being highly correlated. How do we explain this regularity, i.e. the fact that people generally drive on the left side of the road in the UK, etc.? The intuitively appealing answer is that the regularity is *explained* by the existence of the convention of driving on the left-hand side of the road. This would fit both our common usage of the term 'convention' and allow conventions to have explanatory force. Lewis, however, cannot give this type of answer. On Lewis' view the existence of the regularity is *constitutive* of the existence of the convention, and hence cannot explain the existence of the convention. Conventions can only explain regularities if they somehow give rise to them. Such an explanation presupposes that the notion of a convention is independent of that of regularity, and that the existence of the convention is prior to the existence of the regularity.

If we wish to save the idea that conventions explain regularities in action, conventions cannot be equated with regularities. Is there a way of defining the notion of a convention so that conventions to have such explanatory force? One way of doing so would be to restrict the analysis to conventions that exist only in virtue of explicit agreements, or promises to act in a certain way. Take, for instance, a case, where all drivers explicitly agree to drive only on the left-hand side of the road. We can now simply define the notion of a convention in terms of an agreement to follow the agreed upon *rule*. This agreement to follow a certain rule then gives rise to the regularity, and hence we save the idea that conventions explain the later regularity in action.

The above proposal, however, has obvious drawbacks. It can only work for the special case where explicit agreement gives rise to the regularity. Moreover, the whole point of Lewis' work was to show that conventions need not be based on explicit agreements. Fortunately, we can gain the relevant explanatory power of rules without requiring that such rules be followed in virtue of explicit agreements. When people agree to follow a convention this amounts to an agreement to, when a certain set of circumstances occur, act in a certain way. We can distinguish between different ways that it can come about that people follow a rule. One way of making this come about is by explicit agreement, but this is not the only way that it can happen. Such rule-following can emerge spontaneously in any number of ways, most prominently

as a response to a coordination problem. This means that we can identify a convention with a certain *type* of rule, independently of how it came about that the rule is followed.

Note that defining the existence of a convention in terms of a *disposition* to follow a rule also secures the result that a convention can exist even if it has not yet been followed. This is a virtue as, if this was not so, it would *never* make sense to explain the first instance of some action in accord with a convention as being due to the convention in question. Consider, again, the case of two people explicitly agreeing to drive on a certain side of the road. Their agreement to drive on the left uncontroversially constitutes a convention. The first time that either of them drives on the agreed upon side of the road, this act is explained by the existence of the convention explicitly agreed to. But, once again, this can only be so if the convention existed prior to the act of driving. Hence it cannot be constitutive of a convention that it has been followed, but only that the parties to the convention are disposed to follow it.

Note that defining the existence of a convention in terms of a disposition to follow a rule is also useful in explaining, for instance, how a baptism can make it come about that a name conventionally refers to a particular. In a typical baptism, a name is mentioned, not used, and so we cannot say, after a baptism, that the relevant convention has yet been followed. Yet a baptism can make it come about that a person has a certain name, prior to the name actually being used. This is explained by the fact that a baptism can *dispose* people to use a name in a certain way by making some rule governing the use of the name salient. Such a disposition, then, is the fact that is constitutive of the particular having the relevant name.

Lewis does consider the possibility of defining conventions as rules (1969: 100-107). He acknowledges that it is hard to "argue that some conventions are not naturally called rules" (104), but rejects any attempt at characterising conventions as rules. His complaint is "that the class of rules is a miscellany, with many debatable members" (105). This is defended by pointing out the many kinds of things that we call 'rules' that are obviously not conventions.

It is hard to see exactly what Lewis's argument is supposed to be. Lewis argues by constructing a list of things we call 'rules' and pointing out that most of them are not conventions. This, however, is a very weak objection to a view like the coordinating rule view of conventions. On the coordinating rule view it is not the case that all rules are supposed to be conventional, just that some are. One could construct an equally weak argument against the regularity-view of conventions by pointing out that there are all kinds of regularities that are not con-

ventions. This argument would be weak as Lewis is not claiming that all regularities are conventions, but only that regularities of a certain type are conventions. The coordinating rule view similarly only claims that rules *of a certain type* are conventions.

Lewis's also argues that the notion of a rule is "an especially messy cluster concept" (105). This may well be true, but I do not see this as a major objection to characterising conventions as rules in the way I have done. The notion of a rule has been defined here quite precisely as an instruction of the form 'if *C* obtains, perform *K*', where *C* is a situation or context and *K* an action. It takes only a moment's reflection to see that most of what we sometimes call 'rules' (moral norms, rules of etiquette, rules of inference, linguistic conventions, etc.) can be stated in this form, even if the formulation is sometimes a bit inelegant. I do think that uses that do not have this form (i.e. Lewis's example of it being a rule that all meat is more tender if cooked at low temperatures (100)) tend to be examples of loose usage or to be derivative of the notion I have defined here. But, be that as it may, I will not argue that here, as nothing depends on it. How we use the term 'rule' is, ultimately, a matter of mere lexicography, what matters at present is what conventions are. If the reader is unconvinced that my characterization of rules reflects common usage, he can simply interpret my use of 'rule' as a technical term that, by stipulation, applies to injunctions of the form 'if *C* obtains, perform *K*'. Note that nothing of consequence would change if I were to call the view defended here the 'coordinating instruction view', 'coordinating injunction view' or 'coordinating imperative view', or even make up a new term altogether. Whether conventions have the form 'if *C* obtains, perform *K*' is an important matter of substance, whether we should call anything with such a form a 'rule' is, ultimately, trivial.

4.3 Objection 3: Against characterising conventions in terms of exhibited regularities.

Lewis requires that every, or almost every, member of a community must conform to the regularity involved. This is a strange thing to say as presumably it is uncontroversial that a convention can hold in a community even if quite a few members of a community violate it or are entirely unaware of its existence. In later work Lewis clarifies the content of his theory by saying that the idea of a convention which 'holds in a population' can be interpreted in two different ways. It can mean that all (or almost all) of a group follow a convention, or that some sub-population of a population follow a convention (1976, 116). The latter use, then, is the same as my use of the phrase 'sub-group of a population'. On the first usage, he would describe a situation in which there is non-universal conformity, and his other requirements are

only partially met, by saying that in such a case a community has a convention "to a certain degree" (1969: 78 - 80). The basic idea is that, if there is non-universal conformity, the community only has the convention to the degree that the relevant conformity obtains and the other criteria that he lists are met.

On both formulations though, the fact that conventions are defined in terms of exhibited regularities leads to the following oddity: imagine a community of a hundred people in which all are trying to drive on the same side of the road. Stipulate that the community live on a large piece of land and that they do not drive very often at all. In fact, in this community one could drive on the wrong side of the road for quite some time without being alerted to one's mistake, both in virtue of not encountering other cars and, when encountering cars driving on the side different from one's own, assuming that the mistake lies with them. Assume that a convention of driving on the left has been established, either in virtue of explicit agreement or spontaneously in virtue of the context of interaction being a coordination problem. Now imagine that, after some time, five people suffer a cognitive glitch and misremember the content of the convention. They now believe that the convention advises them to drive on the right, and they proceed to do so over an extended period of time. When they encounter someone driving on the left, they simply assume that the other person got it wrong. In such a case it would be uncontroversial to say that there is one convention, namely driving on the left, and that the five deviants are party to the convention, even though they violate it. Lewis, however, cannot portray the situation in this way. Given that Lewis defines conventions in terms of exhibited regularities, he would have to say, on the first usage, that the convention to drive on the left now exists to a lesser degree. Or, on the second usage, that the sub-group within which the convention exists has shrunk. Surely, however, portraying the situation in this way is perverse. We would never say that the convention now only holds to a degree, or that it now exists between less people. Lewis' view departs from common usage[9] and seems to miss something important, namely that the people involved are trying to coordinate, but some are simply failing to do so. The coordinating rule view yields the intuitively compelling answer. All the relevant parties are disposed to, absent defeaters and absent false beliefs, drive on the left. Hence the coordinating rule view yields the answer that there is a convention, namely to drive on the left, and that all one hundred people are party to it. In the same way we

---

[9] The situation may be even worse. Lewis states that the beliefs of the participants need not have general content, but need only be beliefs about the behaviour of specific drivers (1969: 64 – 68). If this is allowed, then the five drivers can have the relevant propositional attitudes required by his theory, and so their behaviour and attitudes constitute a differing convention. On the Lewisian view, then, the situation becomes one where there are *two* conventions, and these exist to radically different degrees.

can then say that, if someone uses 'Quine' to speak about Krugman based on a false belief about the dominant standard governing 'Quine', then this person is still fully party to the relevant convention and fully a member of the relevant linguistic community. The existence of the convention and the size of the relevant community remain unchanged.

Note that the Lewisian analysis would also run into similar trouble where people's behaviour do not exhibit the required conformity due to the occasional presence of internal or external defeaters. On the Lewisian view this would undermine the existence of the convention to some degree. One the coordinating rule view, and here it agrees with our common understanding of the situation, such cases do not affect the existence of the convention. The basic problem with characterising conventions in terms of exhibited regularities is that the required regularities may well not be exhibited, and yet the existence of the convention needs not to be affected in the least. If a German drives on the right in the UK due to a false belief about the dominant standard, or I swerve into an empty right-hand lane to avoid a pothole (external defeater) or to allow a police car to pass (internal defeater), this has zero consequence for whether a convention to drive on the left exists or whether I am party to the convention. This is so, even if the defeaters and false beliefs occur quite frequently. The coordinating rule view captures this.

The three objection discussed above are what motivate rejecting Lewis' definition of convention in favour of the coordinating rule view. Below I will discuss ways in which I think his theory of conventions, i.e. views of his that did not affect the definition of convention itself, should be amended.

5. Objections to Lewis' theory of conventions

5.1 How conventions are sustained

Consider a coordination game in which ten subjects have to name an author. Stipulate that the subjects win a prize if they successfully coordinate, that the amount of money doubles with every added person who names the author that is named by the most people, that the money is split between all participants and that all the subjects have only one concern, namely getting as much money as possible. This is as pure a game of coordination as we can imagine. The context is strategic and all subjects have the same ordinal ranking of all possible outcomes. In such a game there are multiple equilibria, in fact there are as many equilibria as there are au-

thors in the world. If the subjects are allowed to communicate we can expect that rational subjects will agree on an author and all name the same author.

If the subjects are not allowed to communicate we can expect them to try and find some author who is mutually 'salient'. Schelling (1960: 54 - 58) originally pointed out that agents are surprisingly good at solving coordination problems, even if given very little information about the nature of the problem or about the agents they are supposed to coordinate with. Schelling performed experiments in which he gave people problems like meeting in New York City on a specific day, without being told where and when. An absolute majority of subjects coordinated on meeting at Grand Central station at the information booth, and almost all chose 12 noon as the time of the meeting (Schelling, 1960: 550). Schelling explains this surprisingly high level of coordination by saying that subjects find a focal point to coordinate on, where this focal point is a solution that is salient in some particular way (1960:57). The phenomenon of salience is difficult to characterise, but easy enough to recognise. If subjects to a coordination game have a list of six geometric figures to choose from and this list included five highly irregular shapes and one square, the square would be salient. If we give the subjects a square, a triangle, a rectangle, a parallelogram and a complicated, highly irregular shape, then the irregular shape would be salient. If we give the subjects a list of numbers '1, 3, 7, 9, , 2, 5', then we would expect to be picked in virtue of salience, and so on.

We can reasonably assume that subjects will use some criterion of salience as an equilibrium selection rule, i.e. they will do something like trying to identify the author who is such that it is most likely that it is common knowledge that (s)he is salient to the people involved in the experiment. We can, furthermore, be reasonably sure that there will be some authors who are mentioned more than once, i.e. that the odds of an author being named more than once is higher than chance. In fact, intuition suggests - and some informal testing on my students has confirmed – that we can, if the subjects are Westernised English-speakers, be reasonably sure that the author most likely to be picked is Shakespeare. This result would fit well with Schelling's theory, as, if any author qualifies as 'salient' among speakers of English, it is Shakespeare.

Consider a case where we have run such an experiment, that Shakespeare was named by four subjects and that this made Shakespeare the most popular choice. Now imagine that, using the same subjects, we repeat the experiment a hundred times. We would expect dramatic convergence in the answers given by the subjects, i.e. that all, or almost all would name Shakespeare

in the second round, that even the most dim-witted would have caught on by round three and that, by round five, all the subjects will answer 'Shakespeare' for as long as we care to repeat the experiment. In other words, a convention to answer 'Shakespeare' emerges among the participants.

How is this convention sustained? One theory about what happens is to say that, whereas 'Shakespeare' was naturally salient in round one, by round two it is 'salient in virtue of precedence'. Lewis, in fact, viewed salience by precedence as the equilibrium selection rule that explains the stability of conventions over time (1969: 36). The basic idea is that, where conventions are concerned, the equilibria chosen in the future will resemble those chosen in the past precisely because 'having been chosen in the past' makes an equilibrium salient to the future. Note that, in this way, even accidents of the first round can quickly become strongly entrenched conventions.

While I do not doubt that precedence can make an equilibrium salient, I do not think that salience by precedence is what accounts for the stability of our conventions over time. To see why, consider the fiftieth round. If we are to take Lewis' idea of salience by precedence seriously as a theory of how conventions are sustained over time, we have to portray the subjects as, in some sense, achieving coordination by utilising a mental process whereby they try and find an equilibrium selection rule and then choose 'salience by precedence' as the relevant equilibrium selection rule. Note that this is so, independently of how we think about the ontology of mental states. If we are realists about mental states we have to think that this process actually occurs, if we are interpretivists we have to believe that this is the most elegant way of accounting for their behaviour in intentional terms, etc.

The claim that this is the best way of portraying the situation strikes me as hugely implausible. On the coordinating rule view of conventions, however, we can portray the situation differently. Whatever coordination is achieved in the first round will, as Lewis suggests, be a matter of natural salience or accident. It is also plausible to think that Lewis is correct about the second round, i.e. that subjects will pick Shakespeare in the second round as it was made salient by the first round. At some point, however, we can expect the subjects to adopt a rule like 'if asked to name an author by the experimenter, pick Shakespeare'. Furthermore we can expect that, whereas this rule will be followed as a conscious policy at the beginning, it will become internalised until the subjects pick Shakespeare without the problem being given much conscious thought. By the fiftieth round the subjects will follow the rule unthinkingly and out of habit. Nothing that can be described as 'using salience as an equilibrium selection

rule' remains. Note that this should be so for any conceivable rational subject for whom cognitive capacity is a scarce resource. This, at least, is the basic idea behind Herbert Simon's influential idea of 'satisficing' (1997: 295) and the resulting literature concerning cognitive heuristics[10].

It is constitutive of the notion of a rule that it applies to a multiplicity of cases. If a rule is adopted at time *t1*, then following the rule at times *t2*, *t3*, etc. is explained by the adoption of the rule at *t1*. In fact, the whole *point* of adopting a rule is that it provides a subject with a default strategy to follow when some event recurs. Once a rule is adopted, and given that it is not the kind of rule subject to phenomena like *akrasia*, then action in accord with the rule no longer requires explanation. Rather behaviour that deviates from some pre-selected default strategy, i.e. a rule, is what would require explanation. Such an explanation would take the form of identifying a false belief that caused deviation from the rule, or identifying a defeater that caused the rule to be violated in a specific case, though not rejected as a general policy, or identifying a changed circumstance that led to the rule as such being rejected, etc. Acts that conform to a rule that was adopted however, do not need further explanation. The fact that conventions are rules is, by itself, *also* an explanation of the fact that conventions tend to persist. 'Salience by precedence' may well explain why some conventions are adopted, but is not needed to explain the persistence of conventions.

If we characterise conventions as rules this has the twin virtues of, first, being much more psychologically realistic than an explanation in terms of salience by precedence, and, second, explaining the persistence of conventions as deriving from the very nature of, and reason for adopting, rules. There is simply no sense in which my behaviour of driving on the left-hand side of the road every morning, using the euro when in Europe or using 'Quine' when I wish to talk about Quine can be characterised as choosing an option that is salient in virtue of precedence.

5.2 The kind of game that gives rise to conventions

The final topic I wish to discuss concerns how we think of the coordination game which gives rise to conventions. On Lewis' view, the subjects encounter a series of recurrent coordination games. This is correct, but misses two important points. Firstly, once we portray subjects as adopting rules we are portraying subjects as picking a simple default strategy for dealing with

---

[10] See, for instance, Gigerenzer at al. (1999) where it is shown that very simple rules can, under the appropriate circumstances, do as well as, or even outperform, complicated decision-making strategies.

an indefinite amount of future occurrences of a coordination game. What is more, all subjects will expect all other subjects to be choosing such default strategies, all subjects will know that all other subjects are picking such default strategies, and so on. This means that the context of interaction that leads to the adoption of a convention should be viewed as a *meta-coordination game*, i.e. a coordination game about future instances of a coordination game. Given that the future is uncertain, this meta-coordination game is a coordination game *under extreme informational constraints*. When actual instances of the coordination game arise this veil of ignorance is lifted somewhat. This may force subjects to abandon their default strategies in favour of one-off strategies that best serve their interests in concrete cases.

On the coordination rule view of conventions, conventions are best seen as default strategies that arise from meta-coordination games under extreme informational constraints. The second point missed by Lewis – though, in this case, consistent with his view - concerns the distinction between simultaneous move and asynchronous (non-simultaneous move) games. In a simultaneous move coordination game subjects pick strategies without knowing what strategies other subjects will pick, in an asynchronous game at least one subject's strategy is already known to at least one of the other subjects. Note that, in a simple two-person asynchronous coordination game the person who first picks a strategy effectively amounts to a legislator, as the second person, if rational, is forced to coordinate his behaviour with that of the first.

Consider what would happen if we were to, after every third round, introduce a new subject to the Shakespeare game and communication is not allowed. No matter what name the new subject picks in his first round, the other subjects have no incentive to deviate from the rule that they have adopted. What is more, doing so would be costly as abandoning a cognitive habit and acquiring a new one incurs a cognitive cost. The effect will be further strengthened if we stipulate that, as typically will be the case with new parties to a convention, that the older parties are not aware that there is a new party to the convention. This means that, by the second round encountered by the new subject, the new subject effectively faces an asynchronous coordination game as he is simply forced to adopt the rule made evident by play in the first round he encountered. He is, metaphorically speaking, a 'price taker' who simply has to adjust to his circumstances. Note that, unless we suddenly add a lot of new subjects to the Shakespeare game, all new entrants effectively face, in virtue of the older subjects' prior commitment to a rule and informational constraints, such an asynchronous coordination game.

The above reasoning means that Lewis' construal of a convention as sustained by a series of concrete coordination games is incomplete. Conventions are best seen as arising from a meta-coordination game. The adoption of a new convention is best seen as a simultaneous move meta-coordination game, whereas the situation encountered by new parties to a convention will typically be somewhat similar to an asynchronous meta-coordination game. Note that the latter point helps to account for how conventions are sustained despite a constant flow of new entrants to the game. Such new entrants will typically encounter an asynchronous meta-coordination game and can do no better than simply adopting already existent rules. In this way learning that people drive on the left in the UK effectively places the tourist in an asynchronous meta-coordination game, as does learning that Quine is called ''Quine', learning that the euro is used in Europe, etc.

5.3 A first illustration of the relevance of the above to theorising about language

Works on semantics and the reference of names do not typically discuss the nature of conventions in great detail. In this thesis I will argue that such relative neglect of foundational issues has been a mistake. As a first point in favour of the idea that detailed discussions of conventions may be of use in clarifying language-related phenomena, note that the distinction between the meta-coordination coordination game in virtue of which conventional rules come about and the concrete coordination games in which they are employed provides a natural way of drawing the boundary between semantics and pragmatics. Semantics is commonly thought to be a system of context-insensitive rules that are sometimes broken without it being the case that the rule themselves are thereby invalidated. This fits well with viewing semantics as rules that act as default strategies that arise from meta-coordination games under extreme informational constraints. Pragmatics, on the other hand, is typically conceived as the study of the way in which people manage to communicate thoughts by utilising both semantic rules and contingent features of the actual communicative context they find themselves in[11]. These contextually variable features would include things like the common knowledge between a speaker and an audience, their extra-communicative aims, the existence of mutually salient objects, etc. This would fit well with viewing pragmatics as the study of how people act in the actual, concrete games of coordination that the semantic rules pertain to. Furthermore, on such a view we can give a simple statement of the relation between semantics

---

[11] Consider a situation where I communicate that someone is a poor candidate for a job by writing a reference that only states 'He has excellent handwriting'. What is communicated relies on *both* the semantic rules governing the written sentence and contingent features of the situation.

and pragmatics: semantics acts as a set of default strategies that may be overruled, used in a non-standard way, or supplemented in order to communicate if the mutual understanding between the participants is such as to allow this to happen in concrete cases. At the most abstract level, then, the distinction between semantics and pragmatics may well just be a special case of the distinction between rules arising from meta-coordination games and the concrete coordination games that such rules pertain to. I will not try to develop this view any further here. Surely, however, the fact that something like the distinction between semantics and pragmatics can be viewed as deriving from the very nature of conventions, and hence illuminated by the distinction between a meta-coordination game and a coordination game, shows how the tools of game theory can be of help to a greater degree than is commonly supposed.

6. Conclusion

In this chapter I have argued that Lewis' analysis of convention suffers from certain defects that can be resolved by adopting the coordinating rule view. Three of these problems pertain to his definition of 'convention'. These problems are, first, that it requires agents to understand their own situation too well, second, that it robs conventions of explanatory force, and third, that it mischaracterises situations where the relevant behaviour is non-uniform. We can also identify problems in Lewis' theory of convention in general. These problems are, first, that it misses a simple way of explaining the stability of conventions over time and that his characterisation of how conventions are sustained is unrealistic, and, second, he mischaracterises the games that give rise to the existence of conventions. Characterising conventions as coordinating rules, and the existence of conventions in terms of the conditions under which we can say that such rules are followed, allows us to answer these objections and give an intuitively plausible construal of how we should think about the coordination games in virtue of which conventions arise and are sustained.

In the next chapter I will claim that a consideration of the nature of conventions shows that Kripke's causal theory of reference has no non-trivial content.

Chapter 2: The coordination view and the causal theory

1. Introduction

In this chapter I wish to outline a new way of thinking about semantic reference and contrast it with the causal theory of reference. I will start by explaining how semantic reference can be understood in terms of speaker's reference to yield an externalist theory of reference. I will then show that some of the basic properties of such a view of semantic reference derive from the very nature of conventions as such. I will then compare this theory of reference to Kripke's causal theory and claim that Kripke's causal theory has no non-trivial content.

2. Motivation for the coordination view of the semantic reference of names

2.1 Assignment theories and foundational theories of the semantic reference of names

Theories of the semantic reference of names can be understood as having at least two distinct parts. The first part that any such theory must have is a claim to the effect that the semantic reference of names is correlated with some other factor. Call this the *assignment theory*. In this way the traditional descriptivism of Russell (1905, 1910) and Frege (1948) is typically understood as claiming that, for any given name used by a speaker, the semantic referent of the name is identical to the object that uniquely satisfies some descriptive condition that the speaker associates with the name. Kripke's rival causal theory claims that, for any given name used by a speaker, the semantic referent of the name is identical to the object that was baptised with that name at the beginning of the causal chain from which the speaker inherited the name. These claims, interpreted as claims of mere correlation, are to be adjudicated by determining how closely they match how competent speakers would assign referents to names.

We are, of course, interested in much more than merely stating the general form of some correlation between the semantic reference of names and some other factor. Our main interest lies in an explanation of why the claimed correlation holds. One way of discharging the explanatory burden is to give a theory of what semantic reference consists in. Call such a theory, i.e. a theory that attempts to tell us what semantic reference is, a *foundational theory*. The traditional descriptivists, i.e. Russell and Frege, are typically interpreted as having presented such a theory. They are typically interpreted as claiming, in effect[12], that 'being the semantic

---

[12] I follow Kripke (1981) in stating the descriptivist view in terms of reference, despite the fact that Russell would have called the object that satisfies the descriptive condition associated with a name the denotation of the

referent of a name' just *is* a matter of 'being the object that satisfies some descriptive condition associated with a name.' Or, in short, that the reference-relation is a kind of satisfaction relation. The situation with regards to Kripke is somewhat murky. Kripke does not claim to provide a full explanatory theory of why names refer to their referents, or a full foundational theory of what it is for a name to refer to its referent. Yet he is typically understood as having proposed more than the mere correlation claim explained above. Kripke is widely interpreted as having shown that, in some sense, causal chains serve to explain semantic reference.

2.2 The standard view of communication

For the purposes of this thesis it will be useful to have a theory of the semantic reference of names to contrast with the causal theory and traditional descriptivism. In this chapter and the next I will claim that our understanding of the status of both the causal theory and traditional descriptivism is deeply confused. Such a view, as is often the case with critical views in general, is much easier to understand if the reader has an alternative theory in mind that allows for the views under discussion to be compared. *Naming and Necessity* itself provides a good example of how useful it can be to give the reader an alternative picture of some phenomenon when criticising an existing theory of the phenomenon. The bulk of *Naming and Necessity* consists of a series of criticisms of traditional descriptivism. The power of these criticisms, however, is greatly increased by the fact that Kripke also gave a very brief outline of a radically different view of reference, one in which these problems do not arise. Below I will sketch another way of thinking about semantic reference. In fact, the picture that will be provided, even in its embryonic form, is significantly more developed than Kripke's account of the causal view of semantic reference.

In chapter one I developed and defended the coordinating rule view of conventions, which characterizes conventions as rules that we are disposed to follow. As linguistic rules are primarily used to communicate, this means that the coordinating rule view has to be supplemented by a theory of what communication consists in if we are to give an analysis of the notion of conventional reference. The most widely held theory of what communication consists in is a view called the 'conveyancing view', or 'standard view'[13], of communication. On the standard view, communication is an intentional act which consists in a speaker trying to transmit information to a hearer. This information is assumed to be determined by the mental state

---

name. Nothing rests on such matters of terminology.
[13] In calling it the 'standard view', I follow Buchanan (2010). Buchanan is a critic of the standard view, but addressing his criticism falls outside the scope of this thesis.

of the speaker. The hearer's act of interpretation consists in trying to determine what the information is that the speaker is trying to transmit to him. Interpretation, in other words, is a matter of what Carston (2002: 42) has called 'mind-reading'. The standard view, of course, is not without its critics[14], but it is simple, intuitive and widely endorsed[15] among linguists and semanticists. As such it seems as good a view as any to use in order to construct an alternative picture of semantic reference.

2.3 Issues and objections related to the standard view.

In appealing to the standard view I will be assuming the existence of mental states, specifically beliefs and intentions, that have content and contain elements that manage to be 'about' the world. These mental states are supposed to be, in principle, independent of natural language and to exist prior to natural language. I will not attempt to give a characterisation of such content or explain how such states manage to have content. My characterisation of semantic reference will contain an appeal to speaker's reference, about which I will have nothing particularly ambitious to say. In other words, this thesis will be about semantic reference, not about reference as such. The view to be explained assumes that semantic reference should be understood in terms of, and as parasitic on, speaker's reference. This is so, independently of whether we are realists, i.e. think that mental states straightforwardly exist, or interpretivists[16], i.e. think that talk about mental states amount to taking an intentional stance towards certain entities that can profitably be viewed in this manner. Hence I will leave the metaphysics to the metaphysicians and not attempt the resolve the mystery of 'aboutness' as such. Rather the mystery will merely be pushed back.

Two things about proceeding in this way should be noted. Firstly, merely 'pushing back the mystery' is a not a feature that is specific to the view to be explained here. On the standard interpretation of the traditional descriptivism of Russell, the semantic reference of proper names is explained in virtue of an appeal to a relation of satisfaction that holds between elements in the world and descriptive conditions, which must then be analysed in turn. Kripke, similarly, explicitly admits that his causal picture of the semantic reference of proper names

---

[14] See, for instance, Buchanan (2010) and Gauker (1994).

[15] It is endorsed (under different labels) by, for instance, Carston (2002), Åkerman (2010) and Wettstein (1994).

[16] Consider a case where a series of computers transmit information by using a common protocol. If one computer's program for using the protocol contains errors, we won't have any problem taking Dennett's intentional stance and drawing a distinction between the information the computer is trying to transmit, i.e. the information that the corrupted program took as input, and the information it actually transmits, i.e. the standard interpretation of the output of the program. Presumably interpretivists about the attribution of mental states can distinguish speaker's reference from conventional reference in roughly the same way.

does not serve to fully analyse the notion of reference, but presupposes it[17]. Hence my view, inasmuch as it does not analyse reference as such, is on a par with these theories.

Secondly, it is no objection against a view that analyses semantic reference in terms of speaker's reference that it merely pushes the mystery back to another level. There is nothing incoherent in claiming that something which seems to be mysterious in fact not as mysterious as previously thought, as the mystery involved occurs at another level entirely. In this way it is no objection against a theory of economic phenomena that claims that the true mystery occurs at the level of the individual psychologies that produce such economic phenomena that it merely pushes the mystery back a level, provided that the proponent of such a theory does not claim to have resolved the fundamental problem. Similarly, the history of fundamental physics is filled with cases where what is thought to be fundamental is claimed to actually derive from some deeper level of organization that, in many ways, is even more mysterious. The task is to locate the mystery where it actually belongs. This is progress, even if it does not solve all problems at once.

2.4 The objects of coordination and statement of the coordination view

The first issue, if we are to explain the semantic reference of proper names in terms of a co-ordination problem, is to determine what the objects of coordination are, i.e. what is being coordinated with what. In terms of our driving conventions this is simple. Everyone is trying to coordinate the side that they drive on with the side that everyone else is driving on. On the standard view of communication, communication is a matter of a speaker trying to bring some information to the attention of the hearer. If we apply this to names, the most natural thing to say is that, when a speaker uses a name, he is trying to bring some particular to the attention of the hearer. This allows us to define the notion of the speaker's referent:

> A particular *o* is the speaker's referent of an utterance of a name *N* if, and only if, *o* is the particular that the speaker wishes to draw the hearer's attention to by uttering *N*.

There is reason to think that the above may not quite be the correct way to characterise the speaker's referent, but it is close enough for us to proceed[18]. Note that the notion of a

---

[17] Kripke says that the notion of 'reference' is presupposed in his theory's appeal to the notion of 'intending to use the same reference', as well as in the notion of an initial baptism (1981: 97).

[18] The above definition may contain a condition that is only necessary, not sufficient. We may well do better by defining the speaker's referent of a term as the particular which is such that the speaker wishes to make it common knowledge between him and the hearer that he wishes to draw it to the attention of the hearer in virtue of his utterance of a term. I will not pursue this matter any further here as my view only rests on there being some accurate characterisation of speaker's reference.

speaker's referent is an intuitive one that is widely employed. Kripke, for one, has argued that the distinction between speaker's reference and semantic reference is fundamental and applicable to all languages (1977: 267). Nothing in my argument will depend on getting the characterisation exactly right and so I won't get bogged down in detail here.

On the standard view the speaker tries to bring the speaker's referent to the hearer's attention. This means that the relevant objects of coordination are the speaker's referent and the hearer's *beliefs* about the speaker's referent. The basic picture is this: the speaker identifies some particular that he wishes to bring to the attention of the hearer. His task then, is to choose the name that will allow the hearer to determine the particular that he wishes to bring to the hearer's attention. This task has the exact structure of a coordination game. The context is strategic, as the speaker's best strategy depends on what the hearer will deduce, which depends on what the hearer will think that the speaker will think that he will deduce, and so on. Furthermore, the interests of the speaker and the hearer are aligned in that both wish for communication to occur. Hence we can characterise communication that uses proper names as a coordination game in which speaker and hearer are trying to coordinate the speaker's referent with the hearer's belief about the speaker's referent. Note, as indicated in the first chapter, that the coordination game in which rules of usage are chosen should be viewed as being a meta-coordination that provides default strategies to be employed in actual acts of communication.

Coordination games are typically thought to be a matter of coordinating actions. On the surface it may well appear odd to characterise communication as the coordination of the object that an intention is directed at and the object that someone believes an intention is directed at, but I trust that the above explanation shows that there is nothing intrinsically strange about it. Communication, on the standard view, involves a *strategic* context of interaction where the actors' interests are *aligned*. These two conditions are jointly sufficient for a game of pure coordination. Once we have these two conditions in place, there is no reason left to deny that we are dealing with a coordination game, even if it seems slightly odd. Note that Lewis, after only allowing regularities of action to count as coordination games in *Convention*, changed his mind in 'Languages and Language' (1975: 11-12) where he stated that the "proper hearer's response to consider is *believing*" (1975: 11, his italics).

Given that we have defined the notion of a convention, and identified the objects of coordination, we now have all the elements needed to define the notion of conventional reference. I define it as follows:

> A name *N* conventionally refers to a particular *o* in a linguistic community *L* if, and only if, all members of *L* are disposed, absent defeaters and relevant false beliefs, to employ a coordinating rule which advises using *N* to speaker-refer to *o*.

The above definition amounts to a foundational theory of semantic reference. It defines semantic reference in terms of more basic notions, most prominently the notion of a speaker's referent and the notion of a convention. The fact that a certain name conventionally refers to a specific person in a certain linguistic community is portrayed as consisting in the fact that speakers are disposed to follow rules of the form 'if you wish to speaker-refer to particular *o* when speaking to members of community *L*, utter name *N*'. This rule functions as a default strategy to be employed in actual communicative coordination games. As, on this view, semantic reference is essentially a matter of coordination, I will refer to this view as the *coordination view*.

There are various issues concerning the coordination view of conventional reference that naturally arise at this point. One such issue is the relation of the coordination view of the semantic reference of names to historically influential views of the semantic content of sentences that follow essentially the same strategy, i.e. that try to define the semantic content of a sentence in terms of some Lewisian notion of convention and the speaker meaning of a sentence, and then typically try to define the notion of the speaker meaning of a sentence in terms of propositional attitudes, as first suggested by Grice (1957). This issue will only be addressed at the conclusion of this thesis. I will proceed by highlighting various differences between the coordination view and Kripke's causal theory and discussing them in some detail.

Note that I will refer to Kripke's views as a 'theory', even though he claimed to be presenting a 'picture' (1981: 94). I do so as this practice has become quite standard and as nothing in my argument depends on such matters of terminology. The fundamental claim that I wish to defend is that there is nothing in Kripke's positive view about the role of causal chains in explaining semantic reference that is explanatory in the way that it is commonly thought to be. Whether we call this positive view a 'theory' or a 'picture' is of no consequence when determining whether this positive view has any non-trivial explanatory content.

3. The coordination view and the causal theory

3.1 The role of causal chains in the coordination view and in the causal theory

How, on the coordination view, should we think about Kripkean causal chains? To answer this question, consider an iterated version of the Shakespeare game, i.e. where the same sub-

jects are asked to name an author, and this is done repeatedly. Distinguish two variants; in the first the subjects may communicate prior to making their first choice, in the second they may not. We can expect the first game to go as follows: the subjects will agree on an author to name in every round and then all will name that author in every round. Or, at least, something strange would have to happen in order for the participants to not earn the maximum amount of money in every round. This establishes that conventions can originate via explicit agreement. I will refer to this phenomenon as *convention by agreement*.

In the second variant of the game, we can expect the answers given by the subjects to converge quickly. In other words, in the first round only a few subjects will give the same answer, but, after sufficient iterations, the participants will arrive at the same level as achieved by those who were allowed to communicate initially. This establishes that conventions do not need to originate by explicit agreement, but can originate via practice. I will refer to this phenomenon as *convention by practice*.

Stipulate that the Shakespeare game is now complicated by adding new subjects one at a time. If the subjects are allowed to communicate, this does not cause any difficulty. The new subjects can simply be told to answer 'Shakespeare' and so learn the content of a pre-existing convention. From this we learn that conventions can be learned in virtue being explicitly told the content of the convention by those who already know its content.

If communication is not allowed, we can expect the subject to try and guess in the first round, either based on natural salience or not. The subject may get it right, but probably won't. After seeing the results of the first round, however, the subject should be able to figure it out. From this we learn that subjects can learn the content of a convention by being exposed to those who follow the convention. Note that, in both cases, we would expect the subject to be disposed to follow the relevant rule very quickly, and hence to be a full-fledged party to the convention.

Stipulate that, in every third round, a subject already party to the convention is removed from the experiment and replaced by someone else. Soon we should reach a point where none of the original subjects are still part of the experiment. In fact, if we iterate the game a sufficient number of times, we will have a steady sea of changing faces. Soon it will be the case that none of the subjects involved in the experiment were there originally, or even learned the convention from those who were there originally, or learned the convention from someone who learned it from someone who was there originally, and so on. Yet this need not affect the per-

sistence of the convention in the slightest. Provided the changes are not too abrupt the same convention can exist indefinitely.

It has already been pointed out that parties to the convention typically learn the content of the convention from those who know the content of the convention, either in virtue of explicit communication or due to seeing them in action. This has the consequence that everyone learned it from someone who knew the convention, who learned it from someone else, who learned it from someone else, and so on, until we get to the people who learned it from the people who were there when the convention originated. These subjects themselves learned the content of the convention in virtue of being present at the event that caused the convention to come into existence, which could have been either a matter of explicit agreement or practice. Call the event that caused the convention to come into being the *originating event*, and subsequent instances of the convention *downstream conventional rule-following*.

Such instances of learning, of course, do not happen via magic or some non-natural, telepathic process. In any such instance, there will generally be a causal link between the party who learned the content of a convention and the party, or parties, that they learned it from. Simply put, one can rarely learn the content of a convention without being in causal contact with someone who knows the content of the convention. This has nothing to do with conventions as such, but hold for empirical facts in general. The most likely way to learn facts like that a specific table is blue or that the tower of Pisa leans over is in virtue of causal contact with the thing itself, or, in the case of testimonial knowledge, in virtue of causal contact with someone who already knows it. There can, of course, be cases where someone manages to learn something, or at least acquire a true belief, by means of a lucky guess. More importantly, we often deduce truths about specific objects, not in virtue of causal contact with the thing itself, but by our familiarity with objects that are alike in kind. If I know there is rugby match that features Kenya I can be pretty sure that the person playing scrumhalf for Kenya is the shortest Kenyan on the field. But, in general, the knowledge that I have about particulars derives from some sort of causal contact, whether socially mediated or not, that I have had with the thing itself, and this also goes for the events whereby conventions come about.

From the preceding considerations we learn the following, call it the *transmission claim*.

> Transmission claim: Instances of downstream conventional rule-following will typically be causally connected to an originating event.

There can be situations where someone will know that a name has a certain conventional referent, i.e. will know that people are disposed to use a name to speaker-refer to a particular, without this happening in virtue of causal contact with an originating event. If, by whatever means[19], I learn that a certain architect has designed a city, and I happen to know that he always uses the same numbering scheme to name streets and always does the layout in a specific way, I may well be in a position to know that 7th street will be the street with the most shops. This is analogous to how I can know that the scrumhalf is the shortest Kenyan on the field in the example mentioned earlier. Such cases of knowing how a name is used without being in causal contact with an originating event, while unlikely to occur in real life, does imply that we need to formulate the transmission claim in terms of what will typically happen, not in terms of what will always happen.

Note that the transmission claim could not be more trivial. It merely reflects the fact that empirical knowledge about a particular event is typically acquired via causal contact with the relevant event. Also note that the truth of the transmission claim comes about purely in virtue of the fact that causal chains have an *epistemic* role. The role of causal chains in the propagation of a convention lies in the fact that subjects cannot learn an empirical fact, namely that people in a community are disposed to follow a certain rule, via magic. The transmission claim properly understood, cannot be used to suggest that causal chains are somehow constitutively involved in the very nature of conventions as such.

On Kripke's causal theory, a name *N* semantically refer to the particular *o* that was baptised *N* at the beginning of the causal chain from which the user of *N* inherited *N*. On this theory Kripke is committed to the claim that there will always be a causal chain between a baptism and downstream us. Above I have explained why claiming that such a thing is always the case is surely too strong and hence why the coordination theorist should not commit to such a claim. Another difference between Kripkean causalism and the coordination view is that the transmission claim is formulated in terms of originating events as such, and not a *kind* of originating event, namely a baptism. Furthermore the coordination view gives such chains a purely epistemic role, whereas it is not clear what sort of role Kripke thinks that they play. Typically Kripke is read as saying that these chains are the 'mechanism of reference', i.e. that then play some deep explanatory role in a theory of reference. At the end of this chapter I will argue that this idea, while widespread, has no content. I will argue that the view that the Krip-

---

[19] This example adapted from Searle (1983: 239)

kean picture of reference reveals some deep truth about the nature of semantic reference rests on a failure to recognise that the transmission claim is uninteresting.

4.2 How the coordination view and the causal theory handle cases of 'reference-switching'

Returning to the Shakespeare-game, consider ways in which it could come about that the nature of the convention is changed. Let's say the experimenter tells the subjects that they are no longer allowed to use the name 'Shakespeare'. In such a case, if communication is allowed, the subjects can simply agree on a new author. If not, then practice could again allow a new convention to come into existence. Hence conventions can be changed by the same mechanisms that allow them to come into existence, i.e. explicit agreement[20] or practice. Crucially, a change by practice need not come about with the subjects in question being aware that they are changing the convention via repetition. Stipulate that five of the subjects are collaborators, who were told to, at a pre-determined point, start answering 'Milton'. Once it becomes obvious that these subjects are answering 'Milton', and that they do not intend to change, a rational non-collaborator would also start answering 'Milton'. If the subjects are rational, then we will soon have a new convention in place whereby all subjects answer 'Milton'. Furthermore, we do not even need the original five collaborators to have been told to answer 'Milton'. If they suffered some cognitive glitch that caused them to believe that they had been answering 'Milton' all along, the same process will result in all answering 'Milton'. Note that we can make this more realistic and, indeed, inevitable, by tweaking our experiment in various ways. If we change it so that, for instance, some rounds include very few subjects, or that any subject can only learn what answer one other subject gave in a previous round, or that the channel whereby answers are learned becomes 'noisy', i.e. subjects can be mistaken about what someone else said, or new subjects are introduced in large numbers, or we make the game non-simultaneous so that a small number of people can force the hands of others, etc., we can make it inevitable that, sooner or later, the content of the convention will change. The process will be especially sensitive to 'bottlenecks', i.e. cases where the answer of one subject has a large role in determining the answer that will be given by later subjects. Note that, even if, at some later point, all subjects become aware that they are now answering in a way that contravenes the original convention, they may, and it may well be rational for them to do so, choose to stick with the new answer[21]. This is so, especially if the

---

[20] For example, Portugal changed to driving on the right in 1928.

[21] I take it to be uncontroversial that, if all the subjects suddenly come to know the truth about their past behaviour and still do not revert to their original behaviour, it is the case that the content of the convention has changed. We can, of course, describe the situation by saying that a convention went out of existence and a new

new answer is now produced in virtue of an internalised rule that has become habit. Under such conditions there will be a cognitive cost to reverting to the original answer and nothing to be gained from such behaviour.

The positives and negatives of reverting back to an original answer will depend on the specific circumstances and the nature of the convention involved. In general though, we can say the following: conventions can be changed by practice, even if the practice came about by mistake.

Arguably the most famous objection against Kripke's causal theory concerns the existence, as first pointed out in Evans (1982), of cases where the reference of a name changes so that it no longer semantically refers to the object that it referred to at an initial baptism. The most famous example is that of 'Madagascar'. The modern use of 'Madagascar' as referring to the island off the coast of Africa is due to Marco Polo, who, according to most scholars (Oliver, 1977: 219) confused the Somalian port city of Mogadishu with the island off the coast of Africa. Marco Polo's mistaken use spread throughout Europe and so it came about that a phonetically corrupted use of the name 'Mogadishu' now semantically refers to the island off the coast of Africa.

One may reasonably doubt whether we should describe this as a case of reference-switching. This case is complicated by the fact that 'Mogadishu' did not lose its earlier reference, by the fact that the pronunciation was corrupted and by the fact that, at least at the time, the linguistic communities involved were distinct. Personally think that we should describe this as a case of reference-switching. Consider Marco Polo's earliest use of 'Madagascar'. Presumably we would say that, while such use may have speaker-referred to Madagascar, it semantically referred to Mogadishu. Presumably, if he caught his mistake immediately, he would have corrected his usage, as his intention was to convey linguistic information, not create a new usage. In other words, absent false beliefs, he would have been disposed to use 'Madagascar' to speaker-refer to Mogadishu. On the coordination view such a fact is partly constitutive of the fact that his earlier usage of 'Madagascar' conventionally refers to Mogadishu. There came a point, however, where the users of 'Madagascar' would no longer be so disposed. On the coordination view we would express this fact by saying that 'Madagascar' now conventionally

---

one replaced it. I take it that there is merely a verbal difference between these two ways of portraying what happened. Note that we uncontroversially speak of conventions *changing* where one country stops driving on one side of the road and starts driving on another side, as happened in Portugal in 1928. (When a convention 'changes by mistake' what happens is that the events that caused the one convention to go out of existence are the same events that caused the new convention to come into existence.)

refers to Madagascar. Hence I will ignore doubts about whether the 'Madagascar' case is a suitable example. In any case, for present purposes these doubts do no matter. We could simply take a different case, whether real or hypothetical, where reference-switching happened and frame the discussion in terms of such a case.

On the coordination view reference-switching is not in the least strange. It reflects the fact, applicable to any convention, that a convention can be changed by practice, even if this practice comes about by mistake, as has already been demonstrated. Note that the 'Madagascar' case fits well with how the subject of convention change in general was explained in the Shakespeare-game. It has already been mentioned that such changes are most likely in cases where there is a 'bottleneck', i.e. where the usage of a few play a large role in determining future usage. The 'Madagascar'-case is such a bottleneck as Marco Polo was the sole person who made it come about that the name 'Madagascar' was used in Europe.

The coordination view, at least in the case of reference-switching, is superior to Kripke's theory as it gives a natural, non-ad hoc account of such cases. Kripke has never stated exactly how his view should be amended in order to deal with such cases. He has, interestingly enough, mentioned the possibility that one way of dealing with such cases could be to take the view that what is only a speaker's reference at one point may turn into a semantic reference at a later point. He writes that '[w]hat was originally a mere speaker's reference may, if it becomes habitual in a community, evolve into a semantic reference. And this consideration may be *one* of the factors needed to clear up some puzzles in the theory of reference' (1977: 271). On the coordination view this is exactly right. In fact, on the coordination view the idea of 'habitual speaker's reference' can not only help solve some troublesome cases, but also cover all the normal cases. As often happens the supposedly odd case allows the general rule to be seen most clearly.

Kripke, however, has never tried to incorporate the idea behind the above remark into his general 'picture' of reference. In fact, I do not see any natural way for Kripke's view to do so, or deal with cases of reference-switching in general, without either turning the causal theory into the coordination view or, alternatively, rendering the causal theory an analytic truth. This issue will receive further discussion later on in this chapter.

The next issue concerning the relation of the coordination view to Kripke's theory that will be discussed is the matter of how to individuate names.

4.3 How the coordination view and the causal theory handle the individuation of names

The coordination view stands in need of a standard of individuation for proper names. In different contexts words are individuated differently based on our varying interests. In this way an engraver will individuate words by their orthography, a lexicographer will individuate words by their lexical meaning and a logician will individuate words by how they affect logical form. How should names be individuated in order to use the concept of a 'name' in the coordination view? The coordination can only be stated in an elegant form if it turns out that each name has only one bearer. However, if we individuate names by sound or typography, this is false. Furthermore, treating names as having unique referents is at odds with ordinary usage. In everyday conversation we typically say things like 'the most popular name for male babies in 2002 was 'Jonathan'' or 'there are 3 000 people 'Jonathan' in the phone book' and 'several of my friends have the same name'. The only way to make sense of such claims is to interpret the speakers as employing a phonetic standard of individuation.

I do not see any great reason to treat one way of individuating names as correct and the others as somehow improper. Only elegance of expression motivates us to try and individuate names in a way that secures the result that names have unique referents. We could, in principle, formulate the coordination view by individuating names by sound or typography and then treating such names as massively ambiguous. We would need to define some notion like the 'legitimate use' of a phonetic or typographic type and define 'legitimate use' in terms of all members of community, absent false beliefs and defeaters, being disposed to speaker-refer to a particular by using the phonetic or typographic type. It would then be perfectly fine to say that a name, *qua* phonetic or typographic type, has many conventional referents. Nothing of substance would change if we were to do this. It would, however, lead to unnecessarily complicated formulations.

One way to overcome this problem is suggested by Kripke. He mentions, but explicitly states that he does not accept or endorse, the possibility of individuating names by the baptismal event causally responsible for their use (1981: 8n). If two individuals share a name of the same phonetic type, i.e. Napoleon (the statesman) and Napoleon (my dog), then these then count as distinct names as these uses of the name came about in virtue of distinct baptismal events. Note that, even if the coordination view adopted this standard, this would *not* amount to an acceptance of the causal theory. It would be only a convenience, we may well have stated the coordination view by using a phonetic standard of individuation. Kripke himself explicitly states that the adoption of such a standard is merely a convention, and he does not advocate or reject such a convention (1981: 8n). Hence Kripke also views the issue of the in-

dividuation of names as a mere convention[22] which is conceptually distinct from the issue of determining how names refer.

The coordination view, of course, should not adopt such a standard. For a start, the existence of reference-switches would make it the case that one name can have multiple referents, even if we adopted such a standard. Hence the coordination view would, at least, have to formulate the standard in terms of originating events, not baptisms. More importantly, and as has been argued before by using the example of '7th street', someone can use a name to semantically refer even if he is not causally related to an originating event. What is more, someone can even use a future originating event to semantically refer; if a speaker is in possession of the required general knowledge he could make claims about 7th street before the architect really decides to call it '7th street', and before it has been built. It may be objected that such cases are extremely odd – and not to mention far-fetched – but, as has been argued already, the possibility of such cases just rests on the basic fact that I can know things about an individual based on general knowledge about the kind of thing that individual is. The fact that, where names are concerned, this is incredibly unlikely, does not change the fact that there is nothing intrinsically odd about such a form of knowledge-acquisition. Hence I see no non-ideological reason to rule out such cases or treat them as interestingly distinct from the more normal case[23].

It would be best to find a standard of individuation that individuates names in terms of originating events, but without requiring that the use of the name be causally related to the originating event. This can be done once we notice that *any* non-originating use of a name presupposes the existence of some originating event whereby it originated. In typical cases the subject will be causally related to such an event, in atypical cases not. Yet, in both cases there still is some event such that, if an ideally rational version of the utterer of a name with all relevant information had known that it had not occurred, then he would not have uttered the name. If an event is related to an utterance in such a way I will say that the utterance *is premised on* the event. In this way a baptism that caused someone to utter a name counts as an originating event that the utterance is premised on, as does some event whereby a name changed its reference by mistake that then caused a name to be uttered, as does some baptism,

---

[22] Kripke, despite mentioning using the above standard, ultimately states that he will simply adopt a terminology in which phonetically identical names with distinct referents count as distinct names (1981: 7-8).
[23] Devitt (1981) treats such names as attributive, as opposed to referential. The present treatment makes this distinction unnecessary.

or case where a name changed its reference by mistake that the subject is not causally related to, but managed to deduce the existence of, and so on. For present purposes, then, I will individuate names a follows: two utterances of phonetically identical names are tokens of the same type if, and only if, the originating events that the utterances are *premised on* are numerically identical.

The above strategy ensures that all names that semantically refer have unique referents. It remains, however, nothing more than a convenience. We sometimes individuate names in one way, sometimes another, and this serves nothing more than elegance of expression. For present purposes it is convenient to treat each originating event as introducing a new name into our vocabulary. Note that such individuation, i.e. individuation in terms of originating events, would be of no use when giving a detailed explanation of the process of communication. In communication, on the standard view, the hearer is confronted with a token of a phonetic type and has to try and determine the speaker's referent of this token of a phonetic type. Nothing in the process of communication would be helped if he tried to figure out which name, by our standard of individuation, this token of a phonetic type is. The only reasonable way he could do so would be to first determine what the speaker's referent is and then try and determine which originating event such a use is premised on. This will only rarely be possible. Furthermore, it would serve no purpose. As soon as he knows the speaker's referent of the utterance, the aim of his communicative interaction is realised. He has no reason to also try and determine which name, by some irrelevant standard of individuation, he heard. Note that the same reasoning would also apply to using a Kripkean, causal standard of individuation. This standard has nothing to do with communication as, in all but the strangest cases, figuring out what someone is trying to tell us by using a name would be a whole lot easier than trying to determine which baptism the utterance is causally related to, and hence which name, by such a standard, the person has used.

Individuation-by-originating-event is, however, of obvious use to the coordination view. It secures the desirable result that names have unique referents. It also has some use outside of theory as it partly explains how we can, for instance, talk about the different phonetic types equivalent to the English 'Plato' as being the 'same' name, despite the original Greek name for Plato ('Plátōn') being a distinct phonetic type. The uses of these phonetic types are premised on the same originating event, which is part of the reason they can count, for certain purposes, as being the same name.

Note that the 'Plato'/ 'Plátōn' case could also be explained by using Kripke's suggestion of using the baptism that the use of the name is causally related to as a standard of individuation. In fact, using the Kripkean standard in such a case is probably more natural. A further case where the Kripkean standard can be useful is in explaining what we mean when we talk about cases of reference-switching. No-one, including Kripke (1981: 163), has doubted the *datum* that a proper name like 'Madagascar' has undergone a reference-shift. If we individuate a name like 'Madagascar' by phonetic or typographic type it becomes hard to state the content of such a claim. The name 'Madagascar' also refers to other objects, for instance an animated film. Hence we are forced, if we wish to talk of reference shifts in terms of phonetic or typographic type, to talk of distinct uses of names. It is unclear how we are to individuate such uses, if not by some baptism that gave rise to them and the resultant causal chains. It will also not do, in such a context, to individuate names by the originating event that their current use is premised on. If we individuate the name 'Madagascar' in such a way the very concept of a single name changing its referent becomes conceptually impossible. Note that the same problem applies to the standard Kripke does endorse in *Naming and Necessity*, namely simply stating that if, and only if, two phonetic types have the same reference they count as being the same name (1981: 7-8). On such a construal the notion of reference change is also rendered conceptually impossible. Individuation in terms of some baptism, and hence causal chains, offers the only natural standard of individuation on which a claim like 'The name "Madagascar" has shifted its reference' has its intended reading of saying that a single name used to refer to one thing and now refers to another.

As far as I can tell even Kripke would be forced to endorse the 'causally related baptism' - standard to make sense of the idea that one name can change its reference. I see no reason against interpreting such a claim by using the Kripkean standard, or against a general practice of using distinct criteria for individuating names when we wish to do so. We could, of course, express what typically happens in a reference-switch case without using such a standard of individuation. We could say that some particular was named by using a phonetic type and that later on causally related uses of this phonetic type no longer conventionally referred to the particular, but some other particular, and this happened without a new baptism and without anyone involved being initially aware that this has happened. But it is simpler to say that a name changed its conventional reference. As long as we are clear about which standard is being used when, it makes no real difference to anything of substance[24].

---

[24] The notion of a name is not, of course, the only term that we use that presupposes distinct standard of indi-

4.4 The role of mental content in the coordination view and the causal theory

The coordination view straightforwardly makes conventional content dependent on mental content, even if I am agnostic about whether we should be realists about what is attributed by attributions of mental content. Note, however, that the Kripkean theory also makes conventional reference dependent on mental content. First, Kripke's theory semantic reference partially depends on the mental states of utterers, as Kripke requires that users must, when acquiring a name, intend to use it as the preceding user did[25] (1981: 96). Second, note that Kripke's theory depends on the notion of a baptism, which is left unanalysed. Kripke does not give an explanation of baptisms, apart from distinguishing cases where they occur by ostension from cases where they occur by reference-fixing (97). The notion of a baptism may, of course, turn out to be another way in which mental content plays a role in determining semantic reference.

4.5 Rigidity, necessity and semantics

The coordination view implies that names are rigid designators as the speaker's references that are being coordinated are speaker's references to actually existing particulars. In short, if I say that 'Aristotle need not have been named "Aristotle"', I intend to bring the actual Aristotle to the attention of my hearer, not some counterfactual entity that taught Alexander or was named 'Aristotle'. On the coordination view the fact that such usage is standard is constitutive of what it means to say that names are rigid designators. On Kripke's view, the referent of a name is some particular at the end of a causal chain *in the actual world*, not some particular at the end of a causal chain in the world of evaluation. Hence Kripke's theory and the coordination view both naturally lead to the view that names are rigid designators. This means that the coordination view is immune to the so-called 'modal argument' for the same reason that Kripke's theory is immune to such objections.

---

viduation. There are many others; consider 'book'. I can truthfully say that I have forty books on my computer, and also truthfully say that I have fifty books on my book case, but I cannot add these together and say that there are ninety books in my office. E-books and hard copies sometimes count as the same sort of things, sometimes not.

[25] Note that this requirement is too strict. Imagine that I know someone who is doing a survey on names in English, and that I see him walk up to someone and write 'David' on his notepad, or add an extra tick next to 'David'. In such a case the person conducting the survey need not have any intention to use the name at all. His entry on the notepad does not use the name, but, at best, mentions the name. Yet I have now learned that the person if named 'David' and can conventionally refer to him by the name. This, of course, is not problematic in the least if we think of the role of such chains as purely epistemic, as on the coordination view.

Kripke's theory and the coordination view both naturally lead to the view that claims that contain names can be both necessarily true and *a posteriori*. If the individual *o* that we are all disposed to speaker-refer to by using *N* turns out to have some essential property φ, then nothing that does not have property φ can count as being the individual *o*. This is so despite the fact that the discovery that *o* has property φ is an empirical discovery and it being conceivable that it can turn out that we are mistaken, i.e. that *o* can turn out to not have property φ at all. The basic fact is that we can, and typically do, speaker-refer to something without knowing its essential properties. Hence the coordination view is similar to Kripkean causalism in that it forces the view that some claim can be both necessary and *a posteriori*.[26]

Any option available to the causal theory concerning the semantics of proper names is similarly available to the coordination view. In this thesis I will, as Kripke does in *Naming and Necessity,* mainly focus on the theory of reference, and not the semantics of proper names. But it should be clear that any semantic considerations arising from the causal theory should similarly apply to the baptismal view. The most plausible semantics for the causal theory, namely identifying the propositional content of a name with its referent[27], seems equally suited to the coordination view.

4.6 The coordination view and semantic externalism

Semanticists typically distinguish between externalist and internalist theories of semantic reference. On the internalist view semantic reference is supposed to depend only on the mental state of the speaker. On the externalist view, semantic reference depends, at least in part, on social and environmental factors. Traditional descriptivism is typically construed as an internalist theory, Kripkean causalism is a paradigm of an externalist theory.

The coordination view, i.e. the view that a name refers to an individual that the members of a community are disposed to use it to speaker-refer to, is an externalist view. In other words, the mental state of the speaker does not determine what a name conventionally refers to. Rather this is a community-wide affair, and something that an individual speaker can be wrong about. In this way someone may wrongly believe that a linguistic community is disposed to use 'Krugman' to speaker-refer to Lucas, even though the community is disposed to use 'Krugman' to speaker-refer to Krugman. On the coordination view such community-wide

---

[26] Provided, of course, that the things we name have some properties essentially and some accidentally. In this thesis I take no stand on how this distinction is ultimately to be understood.
[27] Kripke himself does not commit to such a view. Soames (2002) has plausibly defended the view that it is the most natural way of providing a semantics that fits with Kripke's views.

dispositions determine conventional reference and so the confused person's own use of 'Krugman' still conventionally refers to Krugman. Hence the coordination view is an externalist theory of conventional reference.

Externalism, however, is not specifically a deep truth about semantics or communication. The content of no interpersonal convention is ever going to depend exclusively on the mental state of an individual following the convention[28]. Interpersonal conventions are a matter of interpersonal coordination; the dominant standard of coordination can, by definition, not conclusively depend on some fact about a specific individual. We should also be externalists about what side of the road to drive on in the UK and externalists about what currency to use in the USA, and so on, as the content of these conventions similarly do not depend on the mental state of the person trying to follow the convention.

Interpersonal conventions are intrinsically social, and hence subject to social externalism. Note that the coordination view, while socially external by definition, can also be external in another sense. Consider a case where the speakers coordinate so that, when they use the name 'Santhon', they intend to bring the first person born in 2014 to the attention of their interlocutor. Such a case, equivalent to what Kripke calls 'reference-fixing', effectively turns the condition whereby conventional reference is fixed into a quasi-indexical. It is then the case that the conventional reference of 'Santhon' depends on the dispositions of the linguistic community, as per usual, but it also depends on certain facts of the real world. In epistemically indistinguishable worlds the phonetic type 'Santhon' could then turn out to conventionally refer to distinct individuals[29]. In this way the coordination view of conventional reference is both socially and environmentally externalist. Note that, even in the case of environmental externalism, we can draw an analogy to conventions as such. Consider a group of adventurers who spend their lives visiting various islands and agree that, whenever they are on an island, they will meet every morning at the spot where their leader first spotted a bird. There is a sense in which, on every island, they will be following a different convention. On one island they will always meet at some specific place, on another island they will always meet at a different specific place, and so on.

---

[28] In fact, even in the case of a strictly one-person convention, like someone who eats at five o'clock every day in order to eat at the same time every day, a qualified version of this claim still holds. Even in such a case the content of the convention will not depend exclusively on the mental state of the person-at-a-specific-time, but rather will be a matter of coordination between time-slices of the person. Such a person could still, at a specific time, suffer a cognitive glitch and be wrong about the content of a convention that only he is party to.

[29] In fact, depending on how we think about baptisms and other originating events, this could be the best way to think of the conventional reference of all proper names.

4.7 The coordination claim and Kripke's objections to descriptivism

It has already been argued that the 'modal argument' does not constitute an objection to the coordination view. Below I will show that none of Kripke's objections to descriptivism pose any threat to the coordination view. Note, however, that I will only treat Kripke's objections as objections to a semantic theory of proper names in a natural language. Kripke is very clear (1981: 25n, 86n) that he is only writing about semantic reference in a natural language. Many, of course, have tried to develop closely related arguments in such a way as to also be applicable to mental content[30]. I do not think that such arguments undermine the notion of 'speaker's reference' that is needed for the coordination view to get off the ground. I will, however, apart from a few suggestions, not address such issues here. Here I will only deal with Kripke's own objections to descriptivist semantics. Note, of course, that Kripke himself has never given any indication that the idea of speaker's reference is somehow suspect, and made unashamed use of the notion in his arguments against Donnellan (Kripke, 1977).

The main objections, apart from the modal argument, that Kripke has against traditional descriptivism all concern the fact that traditional descriptivism places too high a cognitive burden on individual speakers. The first such case is where the description that a speaker may give is not uniquely individuating. Kripke considers the case of a speaker who uses the name 'Feynman', but only knows, of Feynman, that he "is a physicist or something" (1981: 81). But the same can be said of many people, including Gell-Mann. In such a case the description in question obviously fails to do what it is supposed to, namely pick out a unique individual. The second case that Kripke mentions is where the description in question itself includes terms in need of further explanation in order to make it clear that the speaker is picking out a unique individual. If someone identifies Einstein as the man who discovered the theory of relativity, then Einstein is not picked out uniquely until the theory of relativity has been uniquely picked out (82). Kripke suspects that the layman, who is in a position to refer to Einstein, is most likely to explain relativity as 'Einstein's theory' (82). But this would be violate Kripke's condition of non-circularity, and hence fail to identify a unique individual independently. The third way in which the given description can be deficient is that it might not be satisfied by the referent of the name. Here it might be the case that the description refers to someone else, or that it refers to no-one. Kripke considers what would have been the case if Gödel had stolen the incompleteness theorem from a mathematician named 'Schmidt'[31] (84). In such a

---

[30] See, for instance, McGinn (1977: 531 - 535), who applies Putnam's closely related twin earth argument to mental content.
[31] This type of case is not to be confused with the 'modal argument'. Here it is not the case that, in a counterfac-

case the name 'Gödel' would still refer to Gödel, even for speakers who only know, of Gödel, that he discovered the incompleteness theorem. Kripke considers the possibility that we might amend the description 'discoverer of the incompleteness theorem' to something like 'person who published the incompleteness theorem', 'or 'person commonly accredited with discovering the incompleteness theorem', etc., but in each case similar worries can be raised (85). In fact, it seems reasonably clear that a slight change of formulation is not what is called for. These problems can be shown to be more radical by considering, again, the case of 'Einstein'. The layman may well, when asked about Einstein, *identify* him as the 'inventor of the atomic bomb'. No slight change in formulation is going to make this description actually about Einstein. And yet people whose only belief, of Einstein, is that he invented the atomic bomb are still credited with being able to refer to Einstein.

The above objections do not pose any threat to the coordination view. On the coordination view the conventional reference of a name depends on the dispositions to speaker-refer by using a name that exist among the whole linguistic community. The fact that some person may not be able to uniquely identify Feynman does not rule out the possibility that others in his community can speaker-refer to Feynman. The same goes for the case where people are confused, have false beliefs, etc. The social externalism of the coordination renders these objections without force. The objections pose no problem for the same reason that some deluded British driver's belief that it is conventional to drive on the right-hand side in the UK would not undermine the fact that it is conventional to drive on the left-hand side in the UK.

The above objections do, however, lead to a closely related problem. On the coordination view conventional reference is derivative of speaker's reference. This raises the issue of exactly how speaker's reference should be understood, exactly how speaker's reference is determined, and so on. If we assume that speaker's reference must, at least in part, be determined by the beliefs of the speaker, then this problem should be the number one concern of a coordination theorist. Many of the objections to traditional descriptivism, in modified form, are relevant here. Note, however, that the coordination theorist is not nearly as constrained as the traditional descriptivist in trying to respond to such worries. The first important difference is that, if an individual speaker intends to refer to 'whoever is named "Einstein"', then this poses no problem whatsoever. This is so, even if the person's only belief about Einstein is

---

tual world, Schmidt discovers the incompleteness theorem. Rather what happens is that, in Kripkean terminology, it *turns out* that it is Schmidt who discovered it. (In two-dimensionalist terminology, we are considering a possible world as actual, not as counterfactual.)

meta-linguistic, i.e. that he is named 'Einstein'. Kripke objects to traditional descriptivism in the following way:

> If one was determining the referent of a name like 'Glunk' to himself and made the following decision, 'I shall use the term 'Glunk' to refer to the man that I call 'Glunk', this would get one nowhere. One had better have some independent determination of the referent of 'Glunk'. (1981: 72-73).

On the coordination view there is no problem with endorsing the above remark. Indeed, I take it that it is obvious that Kripke is correct. On an externalist semantics like the coordination view, however, there is a matter of fact about the dispositions that members in my linguistic community have, absent defeaters and false beliefs, to speaker-refer to particulars. These facts count as the semantic facts and hence determine the speaker's referent of 'Einstein' of a deluded speaker. Hence even such a speaker can *speaker-refer* to Einstein, due to the simple fact that his beliefs do not determine *conventional reference*. Kripke, of course, has no in principle objection to what he calls 'reference-borrowing' being used to determine semantic reference (1981: 90). There is also no reason to object to speaker's references being, in a sense, 'borrowed' from semantic reference. It is of course, the case that every person in a linguistic community cannot be borrowing his speaker's reference from conventional reference in this way. But the coordination view would agree with this claim. Indeed, on the coordination view such a name would not secure reference, which surely is the correct result.

The first advantage that the coordination view has over traditional descriptivism as a way of dealing with Kripke-inspired objections is that, on the coordination view, conventional reference will not vary if some particular speaker happens to have beliefs that are somehow deficient. The second advantage is that the social externalism of the coordination view allows speakers to unproblematically 'borrow' speaker's reference from conventional reference. In fact, I think that such an idea has considerable intuitive plausibility. I certainly think of myself, when thinking about some person I know almost nothing about, as thinking about 'whoever is named *N'*.

Note, of course, that nothing in this thesis depends on it being the case that speaker's reference should be construed in something like the above way. The coordination view is also consistent with radically different views. The coordination view would not be inconsistent with, for instance, a causal view of *speaker's reference*. Consider a view which states that a partic-

ular *o* is the speaker's referent of an utterance of a name *N* if, and only if, *o* is the particular that caused the speaker to utter N^{32}. The view would, of course, have to constrain the relevant notion of 'cause' in some way so as to answer the *qua* problem, etc. Note, however, that the coordination view would, even in such a case, apply unchanged. It would still make sense to claim that conventional reference is derivative of speaker's reference so that conventional reference is a matter of all the members of linguistic community being disposed to, absent defeaters and false beliefs, use a name *N* when a particular *o* causes the members of a community to utter a name. On this view it would still be the case that conventional reference is essentially a matter of speaker's reference plus game theory, which is the core idea behind the coordination view.

It may be objected that it could be the case that the members of a population may have conflicting dispositions, even absent defeaters and false beliefs. This, however, poses no problem at all and can be dealt with in a Lewisian way (Lewis, 1969). In such a case, as is the case with any number of conventions, we simply have distinct conventions. Some sub-group of a population could have a convention whereby the initial caller calls back if a phone-call is dropped, another sub-group could have convention that the party who received the call calls back, and so on. These conventions could conflict and the conflict could cause problems, but such is the way of the world. In the same manner some sub-groups of society can stubbornly prefer to use the same phonetic type to speaker-refer to distinct particulars, or use distinct names to speaker-refer to the same particular, and refuse to adopt the usage of the other group. This is especially likely where the name has some political or symbolic implication. In fact, something much like this has occurred in post-apartheid South Africa, where the first democratically elected government changed the 'official' names of various places, streets, etc., and where some have refused to adopt new usage, while others have scorned past usage, etc.

The one objection that Kripke makes against traditional descriptivism that does pose an interesting question that the coordination theorist is obliged to answer concerns his famous 'Gödel'/'Schmidt' example (Kripke, 1981: 83 – 85). I take it that it is common cause that, in such a case, the conventional reference of 'Gödel' would remain Gödel, i.e. that we would not have a case like 'Madagascar' where conventional reference would switch. On the coordina-

---

[32] A strict, non-Kripkean causalist who wishes to strip the causal theory of non-reducible mental content would probably have to analyse the notion of a 'baptism' in roughly such a way anyway. In which case he may as well become a coordination theorist.

tion view the basic principle behind reference-switching is easily understood, but it is hard to give an exact statement of the conditions under which conventional reference would switch. The exact statement of such conditions would amount to a statement of the exact equilibrium selection – rule, i.e. the rule for choosing rules, that governs the choice of semantic rules in the meta-coordination game from which linguistic conventions arise. In principle we know that, if the coordination view is correct, then reference-switching would depend on matters of communicative efficiency, i.e. that the criteria that would count are purely pragmatic. It is, however, very difficult to determine exactly what these criteria would be. I will return to this topic at the very end of this thesis, where I will try to show that thinking about the 'Gödel'/'Schmidt' case in game theoretical terms allows us to notice some non-obvious truths about linguistic practice.

5. 'Conventional' reference and 'semantic' reference

I take the terms 'conventional' reference and 'semantic' reference to be largely interchangeable. In other words I take it that, if some particular is the semantic referent of a name, then it is also thereby the conventional referent of the name, and vice versa. This reflects the fact that the semantics of a language is conventionally determined. It is, of course, possible that some philosophers, whether knowingly or not, use the term 'semantic' in a different way. A dispute about whether semantics is conventionally determined would then turn into a verbal dispute concerning the term 'semantic reference', which I take to mean 'conventionally determined reference'. In fact, and this will be argued at length when traditional descriptivism is discussed, I think it is plausible that some the disputes in 'semantic' theorising rests on something very much like a verbal dispute. Furthermore, I will claim that the history of twentieth century theorising about reference has been a history of various conceptual muddles, and, if anything is to be blamed for this, it is the assumption that how we think about 'semantics' amount to a well-posed question.

I have no interest in getting into a verbal dispute about the term 'semantic'. For the purpose of this thesis I will take 'semantic' to mean 'conventionally determined', but nothing of substance will rest on this. What I am interested in, however, is to write about the same topic that Kripke wrote about in *Naming and Necessity*. It is clear Kripke takes himself to be writing about conventionally determined reference. Consider, for instance, the following remarks in 'Speaker's Reference and Semantic Reference' (1977).

> The notion of what words can mean, in the language, is semantical: it is given by the conventions of our language (1977: 263).

> If a speaker has a designator in his idiolect, certain conventions of his idiolect (given various facts about the world) determine the referent in the idiolect[33]: that I call the *semantic referent* of the designator' (1977: 263).

In the above definition Kripke explicitly appeals to the notion of a convention. He also, however, appeals to the notion of an 'idiolect' which, as I will argue in chapter three, is dangerously misleading at best. He does, at least, in two footnotes[34] in *Naming and Necessity*, make it very clear that he is not discussing speaker's reference, but only 'semantic reference'. While it may be tempting to extend his remarks concerning semantic reference of names in a natural language to mental notions, Kripke does not do so in *Naming and Necessity*. Also note that, in 'Speaker's Reference and Semantic Reference', Kripke talks about what words mean, as opposed to what speakers mean, and conceptualises what words mean as matter of a "common language" (1977: 263). Given the above, we may think that Kripke takes himself to be writing about the meaning of names in a public language.

The situation, however, is not quite straightforward. In 'Speakers' Reference and Semantic Reference' characterises the theory in *Naming and Necessity* as saying that the conventions concerning naming include that a given idiolect is no mere idiolect, but forms part of a common language in which names can be passed from link to link (1977: 273). This seems to indicate that Kripke is not *merely* writing about names in a public language, but also claiming that names form part of a public language. In other words, he does not view this as a mere matter of the definition of 'semantic reference'. I take it that this is what is supposed to be the fundamental dispute between him and the traditional descriptivists.

It seems we can interpret Kripke in one of two ways. On interpretation one, he is simply giving a theory of the reference of names in a public language. On interpretation two, he is saying, *contra* what he takes the traditional descriptivists to claim, that our names form part of a public language, and then giving a theory of the reference of such names in a public language. Fortunately for present purposes, on both interpretations he is giving a theory (or, at least, a

---

[33]

[34] These remarks occur in footnotes (p25, footnote 3, reaffirmed in footnote 36) where he briefly discusses Donnellan - style cases of misdescription. "In the text, I speak of the 'referent' of a name to mean the thing named by the name-e.g., Jones, not Smith-even though a speaker may sometimes properly be said to use the name to refer to someone else." (1981: 25n3).

'picture') of how names in a public language get their conventionally determined reference, which is the topic of this chapter.

Note that the above interpretation fits well with the fact that Kripke's work is frequently taken to be part of a general revolution in semantics. Arguably the most important contribution to this revolution, after *Naming and Necessity*, is Kaplan's theory of indexical reference. Kaplan straightforwardly states that "character is set by linguistic conventions" (Kaplan, 1989: 505). Hence, if we wish to interpret Kripke and Kaplan as writing about the same general sort of thing, we have to interpret Kripke as writing about conventionally determined reference.

6. The explanatory content of Kripke's theory

The basic idea behind Kripke's theory of reference (1981: 91) can be stated quite simply. An individual is baptised, either by ostension or in virtue of some reference-fixing definite description. The name is then passed from speaker to speaker, with each speaker acquiring the ability to refer to the baptised individual in virtue of being in such a causal chain. When a 'downstream' speaker utters the name the referent of the name is simply the individual baptised at the beginning of the causal chain. I wish to claim that the standard interpretations of Kripke's work - and probably Kripke himself - ascribe some explanatory content to his work that is just not there. I will proceed as follows. Whatever the explanatory value of the causal theory is supposed to be, it seems clear that it must be more than a statement of certain truisms, as the causal theory is taken to be profound. This claim is often expressed by saying that the Kripkean theory states that causal chains are, in some sense, the 'mechanism of reference'[35]. I don't know what the phrase 'mechanism of reference' is supposed to mean. I will defend the idea that there is no non-trivial interpretation of this claim.

The first truism that Kripke's theory must surely acknowledge, but go beyond, is the *Correlation claim*.

> Correlation claim: The semantic reference of a name *N* is identical to the object that was baptised *N* at the beginning of the causal chain from which the speaker inherited *N*.

---

[35] The claim that Kripke's theory is explanatory and that it concerns the 'mechanism of reference' is widespread. For a typical example, see Reimer's article on 'Reference' in the *SEP*, where she makes frequent use of the phrase 'mechanism of reference' and states that one of the questions addressed by Kripke and rival theories is "*[h]ow* do words refer?" (Reimer, 2009; my italics).

The above should here be interpreted as a claim of mere correlation. It is not explanatory, but the kind of thing that one could discover by generalizing from data about people intuitive judgments about what names semantically refer to. The second truism that Kripke's theory must go beyond, is:

> Causal claim: A speaker can only refer to a particular o by using a name N if there is a causal chain between the speaker and an event where o was baptised N.

The first problem with the above is that the correlation claim, as pointed out by Evans (1982) based on examples like 'Madagascar', is not, strictly speaking, true. I will return to this topic later on in this chapter. For now, let us focus on the cases where the correlation claim and causal claim do hold.

It is implausible to maintain that Kripke and his followers only take him to be asserting the conjunction of these two claims. The correlation claim added to the causal claim does not show that causal chains are the 'mechanism of reference', as the coherence and plausibility of the coordination view attest. More importantly, both claims seem to be no more than shallow truisms about naming. No-one would have been terribly impressed by a sociolinguist who, trying to pre-empt Kripke, gathered data about people's intuitive judgments about the truth-values of various propositions and presented statistics which show that names, even when used in modal contexts, are almost always judged to refer to the person that such names were introduced to refer to. It can hardly be considered a deep, explanatory insight that the vast majority of names, suitably individuated, that are presently in use refer to the individuals that they were introduced to refer to. This just shows that not many cases of accidental reference-change occur. This is mere data that should be accounted for on any theory. Kripke does, in *Naming and Necessity,* present such data, but he is also thought to have put forward a theory (or 'picture) that explains why this data obtains.

The same goes for the causal claim. This reflects nothing greater than the fact that names are public and learned, i.e. that I can rarely obtain semantic beliefs without hearing a name, or seeing it on paper, etc. In fact, it is a trivial fact about all words, i.e. nouns, verbs, adjectives, etc. in a language governed by a public standard that those with semantic beliefs about a specific word will generally stand in a causal relation to the act whereby the word was introduced. Such facts, as applied to names, explain the propagation of *beliefs* about semantic reference and not the propagation of reference itself. In fact, the causal claim is merely an instance of the general 'transmission claim' formulated earlier. No convention can be propag-

ated without it generally being the case that there is some causal chain between the originating event whereby the convention was introduced and downstream, conventional rule-following. Hence neither the correlation claim nor the causal claim explains why names refer to what they do. Ultimately though, the basic problem with attributing to Kripke merely the conjunction of these claims is that he is then portrayed as merely advocating what no reflective ten-year old would deny.

It is, however, hard to say what Kripke's theory asserts, over and above these two claims. Kripke's explanations of his views typically go as follows:

> Someone, let's say, a baby, is born; his parents call him by a certain name. They talk about him to their friends. Other people meet him. Through various sorts of talk the name is spread from link to link as if by a chain. A speaker who is on the far end of this chain, who has heard about, say Richard Feynman, in the market place or elsewhere, may be referring to Richard Feynman even though he can't remember from whom he first heard of Feynman or from whom he ever heard of Feynman. He knows that Feynman is a famous physicist. A certain passage of communication reaching ultimately to the man himself does reach the speaker. He then is referring to Feynman even though he can't identify him uniquely (1981: 91).

There is nothing in the above that is objectionable, except if it is read as saying something interesting about the role of causal chains in *determining* reference. At best it only reflects the truism that causal chains allow a specific language user to acquire the *knowledge* that a certain name can be used to refer. That, however, is not how Kripke is typically understood. Kripke is typically understood as having explained why names have their referents, not how language users learn referring names. Note that the coordination theorist would also endorse the above passage, despite *not* giving causal chains a role in explaining why names refer to the individuals to which they refer.

If we read passages like the one quoted above - and *Naming and Necessity* in general - as explaining how names refer to the individuals that they refer to, then, stripped to its bare essentials, the underlying view goes as follows:

> (1) The user of a name stands in a causal chain which leads all the way back to a baptismal event. (2) The user of a name acquires the ability to refer in virtue of

standing in such a causal chain, as (3) the downstream use of a name refers to the person baptised at this baptismal event. (4) Hence we can see that causal chains, and not identifying descriptions, are the mechanism of reference for names.

The above view mixes triviality with mystery. (1) is a mere truism, which, in slightly modified form, applies to all words subject to public standards. It reflects the fact that the user typically could not have acquired the semantic beliefs that led him to use the name if this had not been the case. (2) is generally true, but we need to be careful about which 'ability' we are talking about. If it is merely claimed that the user needs to be in such a chain in order to acquire semantic beliefs, then this is mostly true, but fairly trivial. If it is the ability to use a name with the required referent that is at issue, then (2) is true, but cries out for an explanation of how causal chains confer this ability. The mere statement of (2) does not amount to such an explanation. This can be seen from the fact that the coordination view also accepts (2), but does so in virtue of the epistemic role of such chains and in virtue of using 'originating events' as a standard of individuation.

Some authors develop claim (2) by saying that downstream users 'inherit the reference' of upstream users, but this is of little help. If they merely mean that downstream and upstream users use a given name with the same reference, this is merely an implication of the correlation claim. If they mean that one needs to be in such a chain in order to acquire semantic beliefs, then this, again, is fairly trivial. If 'inherit the reference' is supposed to mean more, I confess I have no clear grasp of what that would be. The same would go for the claim that upstream users 'pass along' their reference to downstream users. The fundamental *datum* is that upstream and downstream uses of a specific name are co-referential. If saying that reference is 'passed along' is read as merely such a statement of co-referentiality, it is simply the statement of a straightforward implication of the correlation claim. If it is read as saying that semantic beliefs are causally propagated, this is fairly trivial as well. I fail to see any other literal analysis of this phrase.

Claim (3) is generally true, but is merely the correlation claim. It does not serve to explain *why* the name refers to the baptized individual. Given what has come before, the 'hence' in (4) is misguided, as no explanation has been given. Also note that there is nothing in (1) - (3) that justifies the use of the phrase 'mechanism of reference' in (4), or that serves to give an indication of what the phrase means.

Kripke's theory, as portrayed above, does not succeed in giving any kind of answer to the question of why names have the referents that they do. On the above construal it merely states the correlation claim, the causal claim and then treats this as an explanation. I conclude that his view has no non-trivial content.

6.4. Can the Kripkean theory be turned into an explanatory theory?

6.4.1 Does the causal chain stretch back to the baptism or baptised?

In this chapter I have interpreted Kripke's theory as claiming that the relevant causal chain stretches back to an actual baptism, not to the baptised particular. There are some formulations on which we could interpret him as saying that the causal chain stretches back to the baptised thing itself.[36] Such a view would, in essence, include a causal theory of baptisms. On the balance though, I do not think that such an interpretation is correct. Kripke explicitly denies that he is giving an analysis of baptisms. In fact, he states that his theory presupposes the notion of reference twice, both in its appeal to the notion of "intending to use the same reference" and in the notion of an initial baptism (1981: 97). Kripke does not, apart from distinguishing baptisms that occur via ostension from those that occur via reference-fixing by description, have much to say about baptisms. Furthermore, he indicates that, perhaps, ostension can be seen as a matter of reference-fixing by description (1981: 97) and states that descriptivism may be correct as a theory of baptisms[37]. Hence Kripke cannot be ascribed the view that baptisms are to be characterised in causal terms, and the resultant view that semantic reference occurs in virtue of the causal chain stretching back to the baptised.

Imagine, however, a theory that did make the above claim. Would we have an interesting theory if we stated that the semantic referent of a name is the particular at the beginning of the causal chain that caused the causal chain to come about? The nature of this causal chain would have to be specified more closely, of course, as there are any number of people who made it come about that such a causal chain exists. But we'll ignore this problem for now. Also ignore, as we did with Kripke's actual theory, cases where there are no such chains. Just consider cases where such a claim holds true.

---

[36] For example: "Obviously the name is passed on from link to link. But of course not every sort of causal chain reaching from me *to a certain man* will do for me to make a reference." (Kripke, 1981: 93; my italics).

[37] "The case of baptism by ostension can perhaps be subsumed under the description concept also. Thus the primary applicability of the description theory is to cases of initial baptism" (1981: 96).

The problem with such a view, as was the case with Kripke's actual view, is that such chains will tend to exist in virtue of purely practical and epistemic considerations. We typically only baptise some particular if we have come into causal contact with it. This is because we only have reason to name something if we think that it exists, and, for most particulars, we come to learn that they exist via causal contact with them. Hence, much as is the case with the chain of transmission, such chains will tend to exist purely in virtue of their epistemic role[38]. Moreover, the bare fact that such chains will tend to exist could not be less surprising or explanatory. We have no reason to think that their existence tells us anything about semantic reference. Hence, even if Kripke is interpreted as making such a claim, this would not save his positive view from triviality.

6.4.2 The prospects of reductive causalism

One way of turning causalism into an explanatory account would be to offer Kripke's theory as a reductive analysis of reference. This would be the kind of explanation given by those who, inspired by Kripke's work, tried to eliminate, or 'naturalise', the notion of reference by claiming that 'referring' just *is* a matter of standing in a certain kind of causal relationship to an object. On such a view the claim that names refer to the object baptised by that name at the beginning of a causal chain is true due to the fact that standing in such a relationship to an object is what reference actually consists in. Such a claim goes well beyond the truisms explained above and would count as an explanation of the correlation claim.

Kripke, of course, does not make any such claims. He explicitly says that his theory does not eliminate the notion of 'reference' (97), that it may well be impossible to do, and approvingly quotes Bishop Butler's *dictum* that 'everything is what it is, and not another thing' (1981: 94). He also says that the notion of 'reference' is presupposed in his theory twice, namely in its appeal to the notion of "intending to use the same reference" and in the notion of an initial baptism (1981: 97). Hence Kripke cannot be ascribed the explanatory theory that reference just *is* a kind of causal relation[39].

---

[38] Almog (1984) has claimed that such chains do not determine reference, but preserve linguistic meaning (1984: 479 – 482). He does not, however, seem to understand this matter of preserving linguistic meaning as simply a matter of the acquisition of beliefs about conventions, but as concerning a 'presemantic principle' that somehow ties utterances to their linguistic meaning (1984: 482).

[39] Interpreting Kripke as *not* committed to the claim that reference just is a kind of causal relation is quite standard. See, for instance, Speaks (2011). Also note that it is unlikely that this reductive analysis of reference is what people have in mind when they say that Kripke showed that causal chains are the 'mechanism of reference'. We don't, in general, express the claim that phenomenon X turns out to reduce to phenomenon Y by claiming that 'Y is the mechanism of X'. Rather we just say that 'X turns out to be Y' or something similar.

Consider, however, a theory that does claim that semantic (conventional) reference just is a kind of causal relation. The main challenge for such a theory would be to specify the nature of this causal relation without ending up with the claim that this causal relation is the one which exists in virtue of the fact that the relevant particular is the conventional referent of the relevant name. Here is where Evans's examples show their bite. Evans's reference-switch cases show that such a theory cannot be formulated by using the notion of a baptism, as commonly understood. It has to be replaced by some broader notion that applies to all events that I have termed 'originating events'. All conventions, as explained earlier, can come about via explicit agreement or practice, even where this practice arises by mistake. The specification of the causal chain would have to take this into account. Note that such a view cannot use the notion of an 'originating event' as it stands, of course, as this notion is defined as an event that gives rise to a name having a certain conventional referent. It would have to specify the nature of the causal chain without using the notion of semantic (conventional) reference, or the analysis would be trivial.

I do not see any way of specifying such a chain without using the notion of conventional reference itself. Consider, for instance, Devitt's claim that the relevant causal chains must be such as to be 'multiply grounded' (Devitt, 1981)[40], where grounding is understood as some sort of direct, perceptual contact. Let us ignore odd cases like '7th street', where there is no causal contact with the object whatsoever, or the naming of objects that will only exist in future. To see the problem with using the idea of 'multiple grounding' instead of 'originating event', consider a case where grounding happens only once. Imagine a man who sees a woman only once and from afar, and becomes obsessed with her. He names her 'Salome' and discusses her incessantly with his friends. There seems to be no non-ideological reason to deny that their use of 'Salome' conventionally refers. By contrast, consider a case where I see Krugman and somehow come to believe that his name is 'Lucas'. I now use the name 'Lucas' to speaker-refer to him while looking at him. My friends assume that I know what I am talking about and similarly speaker-refer to him by using 'Lucas'. Stipulate that we all look at Krugman while we discuss the contribution that the man we call 'Lucas' made to economic theory. If we were to eventually learn about our prior mistake we could, of course, stick with our usage and use 'Lucas' to conventionally refer to Krugman among ourselves. In such a case we have effectively turned ourselves into a tiny linguistic community where the use of 'Lucas' is concerned. We could also, however, and this is the much more likely course of ac-

---

[40] Devitt (1981), of course, is best interpreted as a hybrid theorist, as he takes the speaker's mental content to supply the sortal needed to deal with the *qua* problem.

tion, drop the use of 'Lucas' and start using 'Krugman' to speaker-refer to Krugman when we learn about our mistake. In such a case we effectively keep the standard of our general linguistic community and 'Lucas' never conventionally referred, even though we successfully communicated by using 'Lucas'. This is so despite there being some causal chain that features Krugman at the beginning of the chain and despite this chain being multiply grounded in Krugman.

In the 'Salome' - case one grounding was sufficient for a convention to be established, while in the 'Lucas' - case, despite numerous groundings, 'Lucas' never conventionally referred. This raises the question as to how many groundings are needed in order for semantic reference to be fixed. This is the core problem that I do not see how the radical causalist can answer in a satisfactory way[41]. The reductive causalist cannot say that the amount of groundings that is needed is the amount required in order for the relevant name to semantically/conventionally refer to the relevant particular, as it does not help to specify the causal chain in terms of semantic reference. Even if we ignore cases like '7th street' where zero groundings are needed, we can still think of cases where one grounding is enough and dream up cases where a hundred groundings are not. The reductive causalist needs some theory of the conditions under which a series of groundings are a 'sufficient amount' of groundings. But this will vary wildly based various contingencies like the prior use of the name, our communicative interests, beliefs and dispositions, etc. It will vary, in the case of typical reference-switches, based on facts that do not yet exist at the time of grounding, as after the grounding there will still be a time where we would still revert to the original use. If our practices are at all guided by practical concerns it can even vary based on facts about phonetics. Consider a case where we mistakenly start calling someone 'Radorea'. We may, when we discover our mistake, decide to stick with it between ourselves, based on the fact that a name so rare won't ever lead to miscommunication between us. We would not have done the same if we had called the person 'John'. We may even stick to 'Radorea' as it is 'cool' in a way that 'John' is not. Consider, in this regard, the case of nicknames. A large part of why attempts at giving people nicknames manage to succeed or fail at establishing a convention has to do with whether they are witty or fitting in some way. This is completely independent of the way in which such a nickname may have been grounded.

---

[41] Devitt (1981) deals with such a cases as cases of 'partial reference'. I see nothing to recommend such an *ad hoc* solution and struggle to make sense of the very idea of *partial* reference. At best such an idea would be a last resort once all else has failed.

I do not see any way to give the conditions under grounding can be deemed sufficient without saying that the groundings must be such as to lead to successful conventional reference, which will not do. I also do not see a reason to try and do so. One must not be misled by the fact that the content of a convention at time *t* will be identical to the content determined at the causally related originating event in virtue of which it has the content that it does. As an 'originating event' is defined as some event that originally makes it come about that a convention has a certain content, this claim is analytic, i.e. if the content of a convention is not identical to the content determined in virtue of the event that brought it about that it has the content that it does, then that event cannot be the event that brought it about that it has the content that it does. The fact that knowledge of a convention is typically causally required, the fact that we typically only name those we causally interact with and the above analytic truth does not combine to signal the start of a promising research program.

8. Conclusion

In this chapter I have outlined a foundational theory or reference, namely the coordination view, explained how it relates to Kripke's view of semantic reference, and argued that Kripke's causal theory has no interesting content. What we take to be the content of Kripke's views depends largely on the role we take him to assign to causal chains. Causal chains can be claimed to explain the propagation of semantic beliefs, to usefully individuate names in some contexts and provide a way of reducing the notion of 'reference' to something else. The first claim is a truism and so, presumably, does not capture the full content of what his views are taken to be. Kripke, furthermore, explicitly states that his theory does not make the last two claims. This leaves it a mystery what he does claim, as I am not aware of any exposition of his work that does not, ultimately, reduce Kripke's views to an endorsement of some mixture of the first these options. But, if that is the full content of his claims about causal chains, then there is no sense in which his theory concerns 'the mechanism of reference' or explains 'how words refer'.

When informally presenting these views I have been met with responses claiming that Kripke is only endorsing the correlation claim, or that Kripke endorses the correlation claim and the claim about the propagation of semantic beliefs, or that he is a reductive causalist, and others. The fact that we are not singing from the same hymn sheet on something so basic - that is generally taken to be common cause - is troubling in itself. I think such disagreement indicates that we are too cavalier in assuming that we know what the semantic relevance of Krip-

kean causal chains could be. Or, for that matter, that we are too quick to believe that they must have some semantic relevance. The basic problem with claiming that causal chains have some deep, explanatory role is that *any* vaguely plausible externalist theory of reference will result in it typically being the case that there is some causal chain between a baptism and downstream usage. On any such theory the referent of a proper name is something to be learned by individual speakers, who can only acquire such knowledge in virtue of causal contact with an initial originating event. Hence the chains exist anyway, in virtue of semantic belief propagation. This means that it is a complete *non-sequitur* to infer, from the fact that such chains exist, that reference is explained by such chains. This can also be seen from the fact that the coordination view is similarly committed to the existence of such causal chains, but need not assign them any semantic explanatory role[42].

How did the idea that Kripke's theory is 'deep' come about? Here I can only speculate. I think that this is partly due to the fact that the philosophical community did not realise that the existence of such chains can be derived from truisms about conventions, coupled with a failure to realise that phrases like 'mechanism of reference' have a clear sense. This is certainly part of why I used to think that the causal theory is profound, and I suspect that others are in the same boat. More importantly though, I think that the confusion arose as Kripke's theory is presented in the context of his arguments against traditional descriptivism. If one is thinking about traditional descriptivism, then presented with Kripke's causal picture, it comes as a revelation. When compared to traditional descriptivism Kripke's theory is so intuitive and so plausible that one cannot but be enticed. What is more, the fact that it is not making any nonobvious claims is then further hidden by the fact that it seems to contradict traditional descriptivism. If Kripke's claims were so obvious, then why did Russell and Frege deny them?

In the next chapter I will claim that Russell and Frege confused semantic reference with the beliefs individuals have about semantic reference.

---

[42] I wish to emphasise that the argument does not depend on the specifics of the coordination view. Any view must have the consequence that there can be succesful baptisms, i.e. baptisms that make it come about that a name conventionally refer to a specified particular. This means that anyone present at such a baptism is in a position to *learn* a semantic fact in virtue of a causal relation. The same goes for anyone present at a non-baptismal, downstream use of the name. Hence, in typical cases, there will always be causal chains that have to be understood as playing a purely epistemic role.

Chapter 3: The coordination view and descriptivism

1. Introduction

This chapter concerns descriptivist theories of semantic reference. In the first part I argue that considerations concerning the nature of conventions allow us to see that the theories of Russell and Frege are false, and that descriptivist views that attempt to go beyond Russell and Frege do not fare much better. In fact, it will be argued that traditional descriptivism is so blatantly false that we need to explain how it ever came about that Russell and Frege held the views that they did. In the second part of the chapter I attempt to provide such an explanation. I will claim that Russell and Frege confused the semantic reference of names with the *beliefs* that people have about the semantic reference of names. I will also argue that such confusion stems from the view that the thought which a speaker associates with a sentence will, if the speaker is a competent user of the sentence, have the same content as the sentence itself.

2. What's wrong with traditional descriptivism

2.1 The problem concerning mistakes

A descriptivist theory of the reference of proper names states that the referent of a name N refers to an object o in virtue of the speaker associating some descriptive condition with N that is uniquely satisfied by o. Such a view is commonly ascribed to Russell and Frege; in Russell's case this 'descriptive condition' is determined by the content of the definite description that the utterer associates with the name (1910: 114), in Frege's case it is determined by the Fregean sense that the utterer associates with the name (1948: 210).

Descriptivism, so construed, is absurd. Consider the Shakespeare game. There is a convention to answer 'Shakespeare' in the Shakespeare game as long as all parties are disposed to, absent false beliefs and defeaters, follow a rule that instructs them to answer 'Shakespeare' when the experimenter asks them to name an author. Note that, at any point, an individual can make a mistake, and come to believe that all the other coordinating parties are disposed to answer 'Milton'. If this mistake spreads and persists, there can come a point at which answering 'Milton' becomes conventional. Until such time, however, such a belief is best characterised as a false belief about an empirical fact. The relevant dispositions to follow the coordinating rule are constitutive of the empirical fact. In this way someone can be wrong about what 'Quine' conventionally refers to, about what side people in the UK drive on, about which currency is used in China, and so on.

The above seems to be the merest common sense, yet Russell and Frege seem committed to denying it. They claim that speakers associate descriptive conditions like 'the first Chancellor of the German empire' (Russell, 1910: 115) and 'the teacher of Alexander the Great who was born in Stagira' (Frege, 1948: 210) with proper names. This seems, at worst, plausible. These descriptive conditions, however, are then taken to *determine* semantic reference. This cannot be right, for it rules out the possibility that these descriptive conditions are simply wrong; i.e. that the speaker associates the wrong descriptive condition with the name.

Kripke's so-called 'epistemic arguments', i.e. cases like 'Gödel'/'Schmidt', 'Feynman'/'Gell-Mann', 'Peano'/'Dedekind', etc. all trade on the descriptive condition relevantly associated by the utterer of the name being somehow deficient; we can make a similar argument by considering garden variety cases of linguistic mistakes. Consider a case where Paul joins a conversation between Bob and John. Bob and John introduce themselves by simply uttering their names, but Paul mistakenly thinks that each is introducing the other. We can grant Russell and Frege that Paul now associates some descriptive condition with 'John' and a distinct descriptive condition with 'Bob'. What seems intuitively clear, however, is that these descriptive conditions do not determine semantic reference; rather Paul now associates the wrong descriptive conditions with each name. 'John' still semantically refers to John and 'Bob' still semantically refers to Bob, despite Paul's mistake.

The basic problem is that on the descriptivist view there is no standard over and above the condition an individual associates with a name, in terms of which the relevant descriptive condition can be judged to be the wrong descriptive condition. The descriptive condition functions, in effect, as a *stipulation* and so there is no possibility of it simply being wrong. This, however, is absurd. While we generally employ names correctly, mistakes do happen, and we simply correct our usage when such mistakes are discovered.

2.2 Three unsuccessful responses to the problem concerning mistakes

At first blush it may seem that we can save Russell and Frege from absurdity by claiming that Paul does not count as a competent user of 'John' or 'Bob', and then claim that traditional descriptivism is only supposed to apply to competent users of a name. The problem with such a response, however, is that on such a construal descriptivism would not be a theory of semantic reference, but presuppose such a theory. The matter of being a competent user of a name amounts to using a name correctly, and such an account of correct usage will, on any vaguely plausible theory, presuppose or imply some theory of the semantic reference of names, i.e. presuppose or imply some commitment concerning the nature of the semantic link

between a name and its referent. This is straightforwardly the case on the coordination view, on which semantic competence would consist in knowing which individual the members of the relevant coordinating community are disposed to, absent false beliefs and internal and external defeaters, use the name to speaker-refer to. Similarly, on the Kripkean view, semantic competence would ultimately consist in knowing who the individual at the beginning of the relevant causal chain of name-transmission is, and so on. Hence we cannot save Russell and Frege from absurdity by stating that their view only deals with competent users of a name, for then the need for a specification of what 'competence' consist in would render the theory circular.

A second possible objection takes its inspiration from the fact that Russell and Frege commit to what I will call *eccentricity* about proper names, the view that the descriptive condition(s) that *the individual utterer* associates with a proper name determines the semantic (conventional) referent of the proper name[43]. It will later be argued that it is this commitment to eccentricity which lies at the heart of the trouble with traditional descriptivism. A defender of Russell and Frege could, however, claim that such eccentricity is not a bug, but a feature, by claiming that traditional descriptivism should be interpreted as the doctrine that our name-employing conventions are personal, and not interpersonal, conventions.

In support of the above idea, note that there is nothing incoherent about the idea of personal conventions. It has already been remarked that a being who has to eat once a day, and has to have these meals as far apart as possible, may well adopt a rule of eating at noon every day. Such a rule is a conventional rule of action. In the same way, an individual can invent a script for recording a secret diary; the rules governing such a script will amount to a set of personal linguistic conventions. In such a convention the individual attempts to coordinate his present and future linguistic behaviour by adopting such standards. Nothing prevents such a script from including names, and so there is nothing incoherent about names that are eccentric, i.e. names governed by the dispositions of a unique individual.

The problem with such a view, however, is that it makes an empirical claim about how we use names, and this claim is false. Denying that there is an interpersonal convention governing 'Quine' or 'Frege' would make about as much sense as denying that the convention of driving on the left-hand side of the road in the UK is an interpersonal convention. For any given in-

---

[43] What I call 'eccentricity' may as well have been called 'individualism'. Burge (1979), however, has already popularised a use of the term 'individualism' that is distinct from what I have defined as 'eccentricity'.

dividual, most of the names employed by the individual are *learnt* from others, not stipulated to be used in a certain way by that individual. We try to follow common standards when using names, we correct our own usage when it is shown to clash with the common standard governing a name, and correct others when their usage clashes with such a common standard. All such behaviour is typical of interpersonal conventions, and the analogous behaviour can be found when we consider practices like driving on the same side of the road, etc. It is not merely a fantastic coincidence that almost all people drive on the left in the UK, rather this is overwhelming evidence of interpersonal coordination. In the same way, it is not a mere coincidence that almost all people use 'Krugman' when speaker-referring to Krugman, 'Obama' when speaker-referring to Obama, and so on. Rather such behaviour is indicative of, and partly constitutive of, the existence of an interpersonal convention. The evidence that the semantic reference of most proper names is determined by interpersonal convention should be taken to be conclusive[44]. Hence we cannot defend eccentricity by claiming that it results from the fact that the conventions governing names are personal conventions.

A third way of defending descriptivism would be to replace the kinds of descriptive conditions offered by Russell and Frege with descriptive conditions that rarely, if ever, render a false verdict as to the semantic reference of a name. Such views do, as explained above, have to respect the fact that speakers sometimes associate the wrong descriptive conditions with a name. But they can easily do so by claiming that speakers associate a number of descriptive conditions with a name, and while some (presumably those similar to those offered by Frege and Russell) can be wrong, the privileged ones, like 'the object at the beginning of the causal chain from which I inherited N', cannot[45]. What is more, the privileged descriptive conditions amount to a standard in virtue of which the other associated descriptive conditions can be incorrect. While it is not clear why such descriptive conditions would be privileged, and somewhat psychologically implausible to suppose that speakers do possess the relevant mental states[46], such views do, at least, have the resources required to deal with the problem concerning mistakes.

We could similarly try to defend Russell and Frege by imputing some sort of cluster theory to them, on which at least some of the relevantly associated descriptive conditions are wrong, as judged against a (weighted) average of the rest of the associated descriptive conditions. The

---

[44] Russell and Frege's views are sometimes referred to as 'famous deeds' descriptivism, i.e. descriptivism that primarily deals with the most famous names. Their examples, in other words, typically concerned exactly those names that are subject to our most enduring *interpersonal* linguistic conventions.
[45] Such causal descriptivism is proposed in Lewis (1984).
[46] This criticism is made by, for instance, Raatikainen (2006).

problem, however, as shown by the 'John'/'Bob' case, is that we can have cases where *all but the most trivial* associated descriptive conditions are wrong. Hence such views inevitably have to, as is the case with causal descriptivism, adopt the view that some suitably trivial associated descriptive condition is the privileged one that determines semantic reference. Note, however, that a view developed in this way would no longer be recognisably Russellian or Fregean, and so can hardly be a defence of *their* views.

Ignoring, however, the fact that such a view would not really be Russellian or Fregean, views on which the semantic reference of a name is identical to some suitably trivial descriptive condition associated with the name by the utterer of the name faces a major objection. Such views, as pointed out by Kripke, run the risk of triviality (1981: 88n)[47]. Suppose the correct theory of reference states that some relation R obtains between a name and its referent. Then, of course, it will be the case that the semantic referent of a name N will satisfy the condition of standing in relation R to N. Furthermore, provided that we can make a case that speakers do associate the condition of standing in relation R to the relevant object with N, it will turn out to be true that the semantic referent of a name is identical to the denotation of some descriptive condition that utterers of a name associate with the name. Such a view *qua* theory of reference, however, is only obviously true if we interpret descriptivism as abandoning all its explanatory ambitions. In other words, while it may be true that the semantic referent of a name is identical to the object determined by some suitable descriptive condition associated with the name by the utterer of the name, this is an entirely different claim from the explanatory claim that a name has its semantic referent *in virtue of* the fact that the speaker associates some descriptive condition with it, or the claim that the descriptive condition relevantly associated by the utterer *determines* the semantic referent of the name[48]. To defend such a claim we would need an argument which not only shows that descriptivism can always render a verdict in accord with our intuitions about what names refer to, but also shows that the relevant descriptive conditions are the ones that bring it about that the name semantically refers to what it does. This is the problem that brings us to the core of the matter, as I will argue below.

2.3 The conclusion we should draw from the problem concerning mistakes

When it was discovered that the candidate descriptive conditions offered by Frege and Russell sometimes do not apply to the semantic referent of some relevant name, some theorists

---

[47] Kripke credits Nozick with this point (1981: 8n).
[48] This point is well made in Kroon (1987).

responded by trying to find some suitably trivial, privileged descriptive condition that will always be true of the semantic referent of the relevant name. This move, however, profoundly misunderstands the source of the trouble. The problem lies with the very idea of eccentricity, i.e. with the idea that the descriptive condition that *the utterer* of the name associates with the name determines its semantic reference. Semantic reference is a matter of convention and, for the vast majority of names we use, such conventions are interpersonal. It is *never* the case, of any interpersonal convention, that the content of the convention is determined by some individual uniquely. It does not matter whether we are dealing with determining the correct side of the road to drive on, or what currency to use in a given area, or the correct application of a proper name, the content of the relevant contention is always a matter of the dispositions that all the members of the relevant coordinating community have to, absent defeaters and relevant false beliefs, follow some coordinating rule. A person may be powerful enough that all other will act in accord with any convention he advises or is manifest through his behaviour. Such obedience though, even if guaranteed, is still constitutively required in order for an interpersonal convention to exist.

If I somehow associate the descriptive condition 'the fastest man in the world' with 'Krugman', 'Krugman' does not suddenly semantically refer to Bolt. The semantic referent of 'Krugman' is determined by the dispositions of the relevant coordinating community; my dispositions have precious little to do with the matter of semantic reference. They factor in, at best, as only one 'vote' in the matter of determining the dispositions of my linguistic community, and hence the semantic referent of 'Krugman'. In the same way, *even when I associate the correct descriptive condition with 'Krugman'*, this descriptive condition, while identifying the correct individual, still does not determine semantic reference. This is why any attempt to find some descriptive condition that all speakers associate with a name and which happens to identify the correct individual is misguided; even if such a condition is found, and even if it can be shown that speakers do associate such a descriptive condition with the relevant name, this still does nothing to explain the semantic reference of the name. Such a descriptive condition, while it may in some roundabout way as one 'vote' among others help to determine semantic reference, does not directly determine semantic reference. This follows from a mere truism about interpersonal conventions.

The argument, then, cannot be simpler. Russell and Frege commit to eccentricity, but this straightforwardly clashes with the fact that almost all of the names we use are governed by interpersonal conventions. The problem concerning mistakes, and the Kripkean cases where the descriptive conditions relevantly associated with a name are somehow deficient, are not

challenges to gerrymander the relevant descriptions into suitable form, but mere symptoms of a deeper underlying problem. This underlying problem is the assumption of eccentricity, which should, as is done on the coordination view, simply be abandoned. If we give up eccentricity, then the problem concerning mistakes disappears, for there is little that is mysterious about the fact that individuals can be wrong about the content of interpersonal conventions.

2.4 Non-eccentric descriptivism?

The coordination view of semantic reference gives up eccentricity and gives up the idea that the convention governing a proper name cites some descriptive condition. Rather it states that names are subject to interpersonal coordination and that the object of coordination is some particular. In other words, it states that the content of the coordinating rule governing 'Krugman' is not 'Use "Krugman" to speaker-refer to the most prominent current Keynesian', but rather 'use "Krugman" to speaker-refer to Krugman'. What, however, are the prospects of a view that gives up eccentricity but still remains descriptivist, i.e. a view that names are governed by rules like 'Use "Krugman" to speaker-refer to the most prominent current Keynesian', or some more trivial descriptive condition?

We can, of course, have a convention that a certain term be used as shorthand for some specific definite description. In other words, abbreviational semantic descriptivism is perfectly coherent. Consider the term 'MVP', used to abbreviate 'most valuable player'. Non-eccentric, abbreviational semantic descriptivism would be the claim that terms like 'John' and 'Toyota' are semantically like 'MVP'. Note, once again, how utterly unlike traditional descriptivism this abbreviational semantic descriptivism would be. The conventional content of 'MVP' is not determined by conditions speakers associate with terms, but by interpersonal coordination concerning 'MVP'; it is something that people can be wrong about and is independent of whatever content some individual may wish to assert by using it.

Based on the above considerations I do not deny that abbreviational semantic descriptivism is intelligible. Note, however, that such explicitly conventional descriptivism is easily refuted as a theory of conventional reference. Firstly, if it were the case that we associate such descriptive conditions with names, one would think that we would be able to state this descriptive condition. In the case of 'MVP', this is easy. Any competent user of 'MVP' knows that it can be used to talk about the most valuable player. No such descriptive condition is forthcoming in the case of case of 'Krugman' or 'Obama'; the descriptivist would have to make the implausible claim that we all do employ the same descriptive condition, we are just somehow

psychologically prevented from becoming aware of what this condition happens to be. Furthermore, we commonly accept that we can be *taught* the content of a name-governing convention by either ostension or a vast set of distinct definite descriptions. The non-eccentric descriptivist would have to, implausibly, argue that such a variety of methods do not actually serve to identify an individual, but actually serve to identify some relevant descriptive condition.

Secondly, note that such a view would be straightforwardly susceptible to Kripke's modal argument. If the defender of non-eccentric descriptivism tries to avoid this problem in the familiar way by rigidifying the relevant descriptive condition, then he is committed to the existence of an extremely odd kind of convention. Such a rigidified descriptive condition would pick out a unique individual in all possible worlds. Yet the non-eccentric descriptivist would have to claim that, even if someone uses the name to speaker-refer to that uniquely identified individual, the person is not a competent user of the name unless the person *also* knows the relevant descriptive condition conventionally tied to the name. Such a convention, even if at all coherent, would be of little use, and there is nothing in our practice to suggest that we have such conventions. As noted above, we allow a variety of ways of identifying an individual to count as learning the linguistic meaning of a name; it would be extremely implausible to claim that such methods actually serve to isolate some rigidified descriptive condition. Furthermore, when someone uses a name and we know that they have the conventionally stipulated person in mind, we do not question them to make sure that they also happen to know what the name actually means. Rather being able to identify the correct person counts as being competent in the use of the name. Hence our name-involving practices suggest that the object of coordination is not some descriptive conditions, but an individual. In this way the convention governing 'George Orwell' is a matter of interpersonal coordination where the object of coordination is George Orwell himself[49]; we try to coordinate so that we all speaker-refer to, and believe each other to speaker-refer to, George Orwell when using 'George Orwell'[50]. Hence I take it that non-eccentric descriptivism, while coherent, is an empirically false[51] theory about the proper names we actually use[52].

---

[49] One useful way of thinking about the linguistic meaning of a term is to think of it as the object of coordination. Hence the linguistic meaning of an indexical is a certain rule, the linguistic meaning of a name is a particular object, etc.

[50] To claim that we coordinate on an individual is not to deny that we need some way of identifying the individual. In fact, it presupposes such ways of identifying individuals, as is explained at the end of the chapter.

[51] Cases like 'Jack the Ripper' are not cases where we coordinate on a description. Rather they are cases where, due to some oddity involving the specific case, we all employ the same proxy-rule in order follow the convention of using the name to speaker-refer to the relevant individual, who remains the object of coordination. The matter

3. Why did Russell and Frege defend eccentric semantic descriptivism?

3.1 Did Russell and Frege defend eccentric semantic descriptivism?

I have argued that descriptivism, in virtue of its commitment to eccentricity, is clearly wrong. This raises the question as to how it came about that Russell, Frege, and their followers, defended a doctrine that is so plainly false. I will attempt to answer this question, but will first pause to consider whether they did, in fact, defend semantic descriptivism. It may seem odd, given that it is the standard view, to defend the view that they were semantic descriptivists. I have three reasons for doing so. First, if descriptivism is plainly false, if I have argued it to be, then we may naturally suspect that they could not have held such a doctrine. Second, Mark Sainsbury has argued, not implausibly, that Russell was not, in fact, a semantic descriptivist. Third, and as should become clear as the discussion progresses, some of the issues that arise when interpreting Russell and Frege help to explain how they ended up defending a position as implausible as semantic descriptivism.

3.1.1 Interpreting Russell

The view that Russell and Frege are semantic descriptivists can reasonably be called the *canonical view*. Russell and Frege have to be construed as writing about semantic content in order for their theories to clash with Kripke's, and, if anything can be called canon, it is the view that there is some substantial clash between Russell, Frege and Kripke concerning the semantics of names. Mark Sainsbury (2002)[53] has claimed, against the canonical view, that the common view of the history of the debate concerning semantic reference is grossly unfair to Russell. The common view is that Russell is a semantic descriptivist in that he believed that the semantic referent of a common name is equivalent to, or abbreviates, a definite description in the mind of the speaker. Sainsbury claims that Russell did have a descriptivist theory, but that his descriptivist theory was not about semantic reference at all (2002: 87). Rather, Russell's views were about 'the thought in the mind of the speaker' (86) upon an occasion of use

---

of 'proxy-rules' is explained at the end of the chapter.

[52] Of course, in the case of names like 'Santa Claus', this will not be possible as there is no such particular. I am tempted to say that it is definitive of a linguistic item 'being a name' that we coordinate on a particular, *if at all possible*. This, then, would not rule out various fall back strategies in cases where, despite our preference for coordinating on a particular, we simply cannot do so. (Such fall back strategies can then include use of stereotypical definite descriptions 'the man who dresses in red and distributes Christmas presents', etc.). In the case of 'MVP' we have no interest in coordinating on a particular, in the case of 'Santa Claus' we presumably would have done so had he existed. Hence such a criterion can still classify 'Santa Claus' as a name, despite it not having a referent. The whole matter of reference to fictional objects, however, is a very difficult matter that I will, beyond the suggestion in this note, not explore any further in the present work.

[53] First published in 1993.

of a name. Sainsbury claims that Russell's interest was in capturing the thoughts and idiolectical meaning of the speaker (89) and not in the semantic referent of a term in a public language, as is the case with Kripke (89). Furthermore, on those rare occasions when Russell does turn his attention to semantic reference, his views are nearly identical to Kripke. Russell, in fact, also views names as Millian, rigid designators (87).

Sainsbury's argument proceeds by way of textual exegesis in support of the claim that Russell was concerned with the thought in the mind of the speaker, followed by an account of how Russell, thusly construed, would deal with certain traditional objections to his theory concerning names. It has to be granted that Sainsbury is correct when he says that Russell has a theory about the thought in the mind of the speaker. Russell, at the start of 'Knowledge by acquaintance and knowledge by description' (1910), states that he is interested in 'what it is that we know in cases where we know propositions about "the so-and-so", without knowing who or what the so-and-so is" (1910: 108). Knowledge is standardly taken to be a matter of belief, i.e. thought, and so Russell is giving a theory of the content of thoughts. Turning his attention names, Russell states that '*the thought in the mind of a person* using a proper name correctly can generally only be expressed explicitly if we replace the proper name by a description' (114; my italics), and states that the relevant description 'will vary for different people' (115). Given such explicit declarations, and others like them, it is indisputable that Russell had a view about thought.

Given the clear verdict presented by the above evidence, it may seem odd that Russell was ever thought to be a descriptivist about the content of names themselves, and not just a descriptivist about the thought in the mind of a speaker uttering a name. The evidence, however, is not as straightforward as the above quotations make it appear. There are, as Sainsbury acknowledges, claims of Russell that seem to indicate that he is talking about the content of words, not thoughts. In each case where Russell made such a claim concerning names, however, Sainsbury shows that the claim is somehow clarified or interpreted by Russell as a claim concerning thoughts. The first example presented by Sainsbury is typical of the general strategy. Consider Russell's claim that 'Common words, even proper names, are usually really descriptions' (1910: 114). Sainsbury objects to interpreting such a claim as a theory about words, not thoughts, on the grounds that Russell immediately clarifies the claim by stating '[t]hat is to say, the thought in the mind of a person using a proper name correctly can generally only be expressed explicitly if we replace the proper name by a description' (114). Sainsbury follows the same strategy in explaining all Russell's claims concerning names that seem to commit him to semantic descriptivism, and makes a good case for his interpretation.

Despite Sainsbury's arguments, I am not convinced. My objection is that, on such an interpretation, Russell seems overly casual about switching between speaking of what names mean and what thoughts speakers who utter names have in their heads. Sainsbury is correct that Russell explains some claims, first stated in terms of what names mean, in terms of thoughts in the mind of the speaker. If, however, Russell never meant to assert semantic descriptivism, it becomes somewhat of a mystery why the statements in terms of what names mean were made in the first place.

The problem becomes considerably more acute when we remember that Russell introduces his view of names in the context of his view about definite descriptions. There is nothing to suggest that these two views are about different topics entirely, i.e. that Russell switches from considering the *semantic content* of definite descriptions to suddenly only discussing the *thought content* of names. Furthermore, when we consider his view of definite descriptions, it is much easier to make the case that he is writing about semantic content. In 'On denoting' (1905), Russell treats the topic of investigation as straightforwardly semantic; he famously states that 'denoting phrases never have any meaning in themselves, but that every proposition in whose verbal expression they occur has a meaning' (480). Russell, throughout 'On denoting', talks in terms of the interpretation of 'phrases', i.e. the meaning of words. Typically, Russell expresses himself by saying things like 'Take as an instance "the father of Charles II was executed". This asserts that there was an *x* who was the father of Charles II and was executed' (481). This claim, and related claims throughout[54], are standardly expressed in terms of some locution indicative of speaking about what words mean, i.e. semantic content.

The first mention Russell makes of names in 'On denoting' occurs in the context of discussing non-denoting expressions. Note that some of his examples of such expressions are names, some definite descriptions; Russell gives examples like 'the round square', 'the even prime number other than 2', 'Hamlet' and 'Apollo' (1905: 491). It would be extremely odd to mix them up in this way if Russell wished to provide a semantics of definite descriptions, but merely a claim about the thought content of names. It is in this context where Russell states that '[a] proposition about Apollo means what we get by substituting what the classical dictionary tells us is meant by Apollo, say "the sun-god'. All propositions in which Apollo occurs are to be interpreted by the above rules for denoting phrases' (491). No suggestion to the

---

[54] Consider, for instance, in 'Knowledge by acquaintance and knowledge by description': 'This conclusion forces us to analyse descriptive *phrases* occurring in *propositions*, and to say that the objects denoted by such *phrases* are not constituents of judgments in which such *phrases* occur (unless these objects are explicitly mentioned)' (1910: 128; my italics).

effect that he is no longer dealing with semantic content is made; in fact, the formulation of the claim in terms of what the proposition (sentence) about Apollo 'means' militates against it[55].

My objection to Sainsbury, then, is as follows. Russell's theory of descriptions is presented as a theory about what certain phrases, i.e. words in a public language, mean. There is nothing to suggest that Russell views his theory of names as being about another topic entirely. Hence, unless we could argue that Russell's theory of descriptions was also only ever about thoughts, we have to conclude that his theory of names is about the semantic content of names, not (merely) the thought content attached to names.

Could we argue that Russell's theory of descriptions was only ever supposed to be about thoughts? I think Russell's formulation of his claims about descriptions as claims about phrases, as contained in propositions (sentences) straightforwardly rules out this interpretation. It is, of course, true that Russell does frequently, most prominently in 'Knowledge by description and knowledge by acquaintance', but also in 'On denoting' and elsewhere, talk about 'what we know', i.e. the contents of thought, when discussing definite descriptions. But he similarly, especially in the case of definite descriptions, also states his claims in terms of what linguistic expressions mean. Russell, in fact, seems to switch between talking about what words mean and the knowledge we can thereby be said to have, i.e. thought content associated with them, *without seeming to think that the different locutions amount to a fundamental change in the topic of investigation*. Examples are scattered throughout his writings, for instance:

> When we say 'the so-and-so exists', we mean that there is just one object which is the so-and-so. The proposition '*a* is the so-and-so' means that *a* has the property so-and-so, and nothing else has (1910: 113).

In the first sentence above, Russell speaks about what 'we mean'. Taken literally, this would be a claim about our intentions, i.e. thought contents. In the next sentence Russell speaks about what the proposition means, i.e. sentence meaning. Such changes in locution seem causal in the extreme if he thought that thought contents and sentence contents are radically different.

---

[55] It will also not do that Russell here means to merely discuss names without bearers as a special class. He nowhere indicates that he views names without bearers as semantically unique. Elsewhere he explicitly dismisses the analogous possibility of treating non-denoting definite descriptions like 'the present King of France' as having different logical form than descriptions that do denote. He states that, based on 'parity of form', they must be treated similarly (1910: 122).

Also:

> When we use the word "Socrates," we are really using a description. Our thought
> may be rendered by some such phrase as, "The Master of Plato," or "The
> philosopher who drank the hemlock," or "The person whom logicians assert to be
> mortal"… (2009: 29).

In the above passage, Russell switches with the same ease between talking of the word 'Socrates' and the thought we have when we utter such a word. Also consider:

> Moreover, the description required to express the thought will vary for different
> people, or for the same person at different times. The only thing constant (so long
> as the name is rightly used) is the object to which the name applies. But so long as
> this remains constant, the particular description involved usually makes no
> difference to the truth or falsehood of the proposition in which the name appears
> (1910: 114).

In the above quotation, Russell starts off by talking about thought content, but, in the last line, talks about the 'proposition', i.e. sentence in which the name appears. Further note that the last line strongly suggests that the content of the *sentence* does vary based on the relevant description, despite the fact that the truth value of the sentence remains unaffected.

The three passages quoted above are not exceptions, but illustrative of Russell's general practice of causally switching between talking of thought contents and semantic content. I can see only one way to account for the ease with which Russell switches, both in the case of definite descriptions and names, between talking about semantic content and thought content. This makes sense if we suppose that Russell thought that, at least in the case of competent speakers, *semantic content and thought content coincide*, i.e. if we can interpret Russell as assuming that, in the case of a competent speaker, the thought content that guides the speaker's act of using a sentence has the same content as the sentence itself. On such a view the speaker thinks that p, wishes to communicate that p and, being competent, sincere and so on, assertively uses a sentence with the conventional content p.

Call the view that semantic content and thought content of utterances, in the case of competent and sincere speakers, coincides in this manner, the view that the relevant utterances exhibits *cognitive-semantic* identity. Sentences exhibit such cognitive-semantic identity if suitably

competent and sincere speakers will utter them if, and only if, they are guided by a thought with the same content as that of the uttered sentence. On the assumption that thought and semantics are compositional in a similar way, we can speak of sub-sentential expressions as exhibiting cognitive-semantic identity in an analogous way.

If a writer assumes that all of natural language assertions exhibit cognitive-semantic identity then such an author can afford to be quite casual about switching between talking about the content of a thought and the content of an utterance. In such a case, what is asserted about thought content is *ipso facto* also taken to be asserted about semantic content, and what is asserted about semantic content, is thereby also taken to be asserted about thought content. If we interpret Russell this way, then his way of articulating his views is no longer almost fantastically sloppy, but merely a matter of ignoring a difference that, given such an assumption, makes little difference[56].

Note that there need not be a clash between the doctrine of cognitive-semantic identity and eccentricity about names (or any other linguistic item). There may seem to be a problem in that cognitive-semantic identity is only claimed to hold for *competent* users of a name. Yet, if an author commits to eccentricity about names, then the descriptive condition that they attach to a name becomes the only authority relevant to determining competence in the use of a name. This, then, is the problem that lies at the core of Kripke's epistemic arguments against descriptivism and what I have termed 'the problem concerning mistakes'. It would, furthermore, seem to trivialise the very notion of competence. Yet, someone can still maintain a doctrine of general cognitive-semantic identity about linguistic items, even if he views some terms as eccentric, as in the case of eccentric terms, the doctrine of cognitive-semantic identity can still be true, if trivially so.

Ascribing to Russell the assumption that assertions in natural language exhibit cognitive-semantic identity has two virtues. First, it would explain his practice of causally switching between talking of language and talking of thought. Hence it makes sense of Sainsbury's point that Russell often explains or clarifies a remark which seems to be about semantics with a remark which seems to be about thought, but without forcing us to agree that Russell only ever intended to speak about thought. The second virtue is that, on this interpretation, we do

---

[56] Note that this would also make sense of his claim that 'On denoting', while much more focussed on semantic content, and 'Knowledge by acquaintance and knowledge by description', while much more focussed on thoughts, are dealing with the same topic (1910: 108).

not need to convict the vast majority of the profession of a systematic interpretation of Russell. It would vindicate Kripke, and the canonical interpretation of Russell. Consider a claim like the following:

> The word 'German' will again have different meanings for different people (1910: 115).

The above claim can be interpreted in its evident sense, namely that the semantic content of a name depends on who utters it, and not as merely being a clumsily expressed claim about thoughts. This is so, even if we do take Russell to also be committed to a claim about varying thought contents.

What are we to make of Sainsbury's claims that Russell is a Millian? This interpretation is based on the fact that Russell portrays communication as only occurring in virtue of the fact that, when a name is asserted, there is some singular proposition, known by description, that the speaker and hearer share. The passage which most strongly supports this contention is the following:

> It would seem that, when we make a statement about something only known by description, we often *intend* to make our statement, not in the form involving the description, but about the actual thing described. That is to say, when we say anything about Bismarck, we should like, if we could, to make the judgement which Bismarck alone can make, namely, the judgement of which he himself is a constituent. In this we are necessarily defeated, since the actual Bismarck is unknown to us. But we know that there is an object B, called Bismarck, and that B was an astute diplomatist. We can thus *describe* the proposition we should like to affirm, namely, 'B was an astute diplomat', where B is the object which was Bismarck. If we are describing Bismarck as 'the first Chancellor of the German Empire', the proposition we should like to affirm may be described as 'the proposition asserting, concerning the actual object which was the first Chancellor of the German Empire, that this object an astute diplomatist'. What enables us to communicate in spite of the varying descriptions we employ is that we know there is a true proposition concerning the actual Bismarck, and that however we may vary the description (so long as the description is correct) the proposition described is still the same. This proposition, which is described and is known to be true, is what interests us; but we are not acquainted with the proposition itself, and do not know *it*, though we know it is true (1910: 116).

On Sainsbury's interpretation of the above passage, Russell commits to the claim that, while our thought contents may include some descriptive proposition, the sentence content is some singular proposition concerning Bismarck himself. It is a point in favour of his interpretation that such a view is much more plausible theory of sentence content than the view that results from adopting (eccentric) semantic descriptivism about names. The question, however, is not which theory is more plausible, but which theory Russell actually held. And, for reasons already stated, I think the evidence strongly suggests that Russell held the more implausible view. The evidence, as already explained is that, firstly, Russell's theory of descriptions is about semantic content, and Russell never indicates that his descriptivist view of names is about another topic entirely, and, secondly, that Russell seems almost indifferent between stating his views in terms of thought content or semantic content. Such casualness would be almost criminal if Russell truly believed that thought content and semantic content come apart. Furthermore, whereas Russell does present, in the Bismarck case, the relevant singular proposition as necessary for communication, he *nowhere* suggests that he thinks that this singular proposition is the semantic content of the sentence which includes 'Bismarck'. In fact, the above passage militates against such an interpretation; in the first sentence Russell says that we would very much like our statement, i.e. *sentence*, to be about Bismarck, but then proceeds to explain why this is *not* the case. I cannot see how such a claim can be interpreted as not being a denial of the view that the *sentence* is about Bismarck. Hence, while Russell does mention the singular proposition that the Millian would take to be the semantic content of the sentence, the evidence suggests that he did not adopt Millianism. Rather it suggests that he took the relevant singular proposition to be somehow *derivative* of the semantic content, i.e. that the semantic content of the same sentence, used by different people to express different semantic content, must *overlap* in determining the same singular proposition in order for communication to occur. Also note that, so construed, the above quotation, and Russell's remark about the word 'German', shows a clear commitment to eccentricity.

There is also further textual evidence that indicates that Russell did reject Millianism about semantic content. In Russell (1910: 123 – 127), he argues against the view that identity claims like 'Scott is the author of Waverley' assert identity of denotation. Such a dispute is about semantic content, if anything is, and for once Russell mostly states his view in terms of semantic content, not thought content. Russell writes as if Millianism were true, and argues that, on such a view, there is no one relation called 'denotation' that holds between 'Scott' and Scott and between 'the author of Waverley' and Scott. The first relation is conventional, the second is factual. I will not judge the argument here; what is important for present purposes is

that Russell explicitly states that he is adopting Millianism *for the sake of the argument* when he says "I neglect the fact, considered above, that proper names, as a rule, really stand for descriptions' (1910: 123). Given that the debate is clearly about semantic content, I fail to see how such a claim is anything but reminder of his denial of Millianism.

The last bit of textual evidence is from *The Philosophy of Logical Atomism* (2009)[57]. Russell states:

> Proper Names = words for particulars… I have put that down although, as far as *common language* goes, it is obviously false…. What pass for names in language, like "Socrates," "Plato," and so forth, were originally intended to fulfil this function of standing for particulars, and we do accept, in ordinary daily life, as particulars all sorts of things that really are not so. The names that we commonly use, like "Socrates," are really abbreviations for descriptions; not only that, but what they describe are not particulars but complicated systems of classes or series. (2009: 28 - 29; my italics)

In the above Russell states that Millianism, *as a doctrine about our common language*, is false, that descriptivism is the correct view of such a matter and that particulars are not the contents of the relevant descriptions. Also note that, in the same passage, he also states that:

> A name, in the narrow logical sense of a word whose meaning is a particular, can only be applied to a particular with which the speaker is acquainted, because you cannot name anything you are not acquainted with (2009: 29).

The above passage strongly indicates that Russell think that epistemological issues strongly determine logical ones, i.e. that what can be the content of a name depends on what can be said to be the content of the relevant thought. This, again, counts in favour of the interpretation I have been urging *contra* Sainsbury, namely that Russell takes thought content and semantic content to coincide, and is discussing *both*.

There is one subtlety concerning Russell's view of thought content that should be noted. Such thought content cannot straightforwardly be equated with the state of mind of the utterer. Russell explicitly states, in his reply to Strawson (1950), that he is not trying to capture that

---

[57] First published 1918.

state of mind of the utterer of a sentence, rather he 'was trying to find a more accurate and analysed thought to replace the somewhat confused thought which most people at most times have in their heads' (1957: 388). Russell makes this point in the context of someone who says 'I ain't never done no harm to no-one' to convey that she has never hurt anyone. Concerning this example, Russell seems to hold that what the speaker means to assert coincides with what she does assert. This fits well enough with the view defended here, namely that Russell thinks that ordinarily thought content and semantic content coincides, provided we do not straightforwardly equate the thought the speaker tries to convey with her exact state of mind[58]. Hence, strictly speaking, we should interpret Russell's notion of a 'thought' as some sort of improved rendering of what speakers actually have in mind. This issue, however, is orthogonal to the current discussion; I will ignore the matter of exactly how the relevant thought content is supposed to be determined.

3.1.2 Interpreting Frege

I take it that it is established that Russell did commit to a descriptivist view about the semantic content of proper names. In the case of Frege, fortunately, things are much more straightforward. Frege does not present his theory as a theory of thought content, but in terms of the content of linguistic entities; he explicitly talks of senses as 'connected with a sign (name, combination of words, letters)' (1948: 208) and states that distinct proper names like 'evening star' and 'morning star' have distinct senses (208). Such senses are 'grasped by everybody who is sufficiently familiar with the language' (208), i.e. they are aspects of a language that are learned by individual speakers. Frege also speaks of the 'customary sense' of a word (212) and, in his 'Letter to Jourdain' (1993: 44), explicitly portrays the explorers who see the snow-capped mountain as learning the co-referential names 'Aphla' and 'Ateb' from local linguistic communities. Such formulations make it clear that he (primarily) presents a theory of semantic content, not some other *species* of content.

The above formulations, however, seem to indicate non-eccentric, i.e. non-individual descriptivism. Frege, however, in his famous footnote concerning 'Aristotle' states that

---

[58] One difficulty, however, concerns the chosen sentence, which does not standardly mean what the person tried to convey. Does the passage indicate that Russell sees an important distinction between sentence meaning and what is asserted? One option would be to say that Russell does simply equate 'what is asserted' with 'what the sentence means'. It is after all, the case that, while the use of the sentence by the person is not standard, there is a conventionalised use of the phrase on which it means exactly what she tried to assert. There is some evidence for this in the text in the text; Russell contrasts the speaker's utterance with 'the rules of syntax which Mr. Strawson would adopt in his own speech'(1957: 388). I am not, however, quite sure if this is correct.

'opinions as to the sense may differ' (1948: 210). This does not, by itself, constitute any withdrawal from full-blown non-eccentric descriptivism. Opinions as to the meaning of any term may differ within a linguistic community, and in such cases some of the users of the term are simply wrong. Frege though, does not go the route of claiming that some name-users are simply wrong, rather such variation is something we can put up with in an imperfect language. Elsewhere he indicates that people whose usage do differ in such a way should be treated as speaking different languages (1956: 297).

There are two ways that we can interpret the above claims. One option would be to interpret it, not as a commitment to eccentric descriptivism, but as a claim that, when it comes to actual proper names[59], the relevant linguistic communities are much smaller than we ordinarily suppose. On such a view, names should be individuated by senses, so that all those people who use the same phonetic type with the same sense can be treated as using the same name. Such a doctrine would not be eccentric, rather it would just shrink the relevant linguistic community dramatically. On such a view, a use of a proper name would not refer in virtue of the sense the speaker attaches to the name, but in virtue of the sense being the sense that the community, albeit small, attaches to the name. All the arguments already made against non-eccentric descriptivism would, of course count against such a view.

On balance, however, I will not follow the above interpretation, but rather interpret Frege, as Kripke does, as an eccentric descriptivist. If non-eccentric descriptivism is to mean anything at all, it must mean that there is some extra-individual standard in term of which a speaker can attach the wrong sense to a name, i.e. the speaker can attach a sense to a name that does determine the same referent as the sense relevantly attached by others, but it is still the wrong sense in being a different sense from that attached by others. The textual evidence, however, indicates that, in such a case, Frege would treat the speaker, not as violating the standards of some small linguistic community, but as speaking a distinct language. On such a view, then, the sense the speaker attaches to the name does determine the content of the Fregean 'thought' expressed by the sentence in which the name occurred; the fact that a number of people attach the same sense to the name does not form part of the *explanation* as to how the Fregean thought expressed by the name comes to have its content. In this way the sense that the

---

[59] As opposed to Fregean 'egeinnamen', which includes what we ordinarily call proper names and definite descriptions (1948: 210).

speaker attaches to the name has somewhat of the character of an individually determined stipulation, and so I will interpret Frege as an eccentric descriptivist.

There is further textual evidence that Frege committed to eccentricity. Consider his discussion of Dr. Gustav Lauben:

> '[I]f both Leo Peter and Rudoph Lingens understand by 'Dr. Lauben' the doctor who lives as the only doctor in a house known to both of them, then they both understand the the sentence "Dr. Gustav Lauben has been wounded" in the same way, they associate the same thought with it. But it is also possible that Rudolph Lingens does not Dr. Lauben personally and does not know that he is the very Dr. Lauben who recently said "I have been wounded". In this case Rudolph Lingens cannot know that the same thing is in question. I say, therefore, in this case: the thought which Leo Peter expresses is not the same as that which Dr. Lauben uttered' (1956: 297).

The above passage indicates that Frege thought that the content of the relevant sentence (i.e. the thought expressed) is determined by how the utterer understands the sentence, i.e. by the descriptive condition *eccentrically* associated with it.

Note, furthermore, that the above passage also strongly suggests the view earlier attributed to Russell, namely that, in the case of competent speakers of a language, the thought content of the sentence expressed coincides with the semantic content of the sentence. Frege seems to move, as Russell did, from epistemological concerns, i.e. concerns regarding what the speaker knows or understands, to logical and semantic ones. In the first sentence he considers thought-content, i.e. the issue of how the speaker understand a sentence and the issue of what a speaker knows concerning an individual. Reasoning from such premises, in the last sentence he draws a conclusion concerning the content (the Fregean 'thought') of the sentence uttered. Such an argument makes sense if one assumes that, under suitable conditions, the content of the thought in the mind of the speaker and the content of the sentence coincides. Strictly speaking, of course, the passage only shows that Frege thinks that thought-contents strongly constrain semantic content. I can see nothing, however, to block the interpretation that he believed, as Russell did, that linguistic expressions generally exhibit cognitive-semantic identity, i.e. that the thought content of the speaker is identical to, and determines, the semantic content.

Above I have explained why I follow the canonical view on which Russell and Frege are interpreted as semantic descriptivists. I have also explained why I view both as committed to eccentricity about proper names, and hope to have made a *prima facie* case that both held the view that, at least in the case of competent speakers, thought contents and semantic contents coincide. This, then, as I will argue later, allows for an elegant diagnosis of how it came about that they held the doctrines that they did.

3.2 What Russell and Frege were *really* thinking about

3.2.1 The notion of 'actually' thinking about something

I have argued that descriptivism, in virtue of its commitment to eccentricity, is clearly wrong. This raises the question as to how it came about that Russell, Frege, and their followers defended a doctrine that is so plainly false. My answer will have two parts. In the first part, I will claim that they confused a subject's belief about the content of a name-governing convention with the content of the convention itself. In the second part I will present my view as to how they made such a mistake. In defence of the first contention, I will claim that, while they took themselves to be thinking about conventionally determined content, they actually, as far as names are concerned, ended up thinking about something else entirely, namely a subject's beliefs about conventionally determined content. When I say that they took themselves to be thinking about one thing, but *actually* thought about something else, I mean that a situation like the following came about.

Consider the situation of Kenneth, who is spying on Simon's movements. Suppose that Simon disappears from view for a moment, and then Fred, who Kenneth mistakenly takes to be Simon, emerges. Kenneth keeps watching Fred and, based on his behaviour, acquires false beliefs about Simon. The situation is then reversed, with Fred disappearing and Simon emerging. Kenneth then goes back to acquiring true beliefs about Simon.

If we were to quiz Kenneth about Simon, we would expect to find that most of his beliefs about Simon were true, but that a small number of these beliefs are wildly false. We can explain this by saying that Kenneth *confused Simon with Fred*. There is a sense in which, even though Kenneth takes these beliefs to be about Simon, they are *actually* about Fred.

It is quite tricky to explain what it is for a belief to be 'really about' some topic, despite the person involved thinking that they are about something else. Such attributions are perfectly

natural, but hard to characterize exactly. Maybe we should express what this sense of 'about' amounts to by saying that Fred was the direct cause of the relevant beliefs and that 'aboutness' here just amounts to 'being caused by'. However, while causation may be a necessary feature of this sense of 'aboutness', it is far from sufficient. If I come to hold that pencils are green, purely due to a knock on the head or as a result of ingesting some mind-altering substances, we would not say that my belief is 'really about' the knock on my head or about the mind-altering substances. The causation that is involved in such cases did not come about in the appropriate way. Perhaps we should modify this by saying that the person thinks his thoughts are caused by one thing, whereas they are actually caused by another thing. However, independently of whether we can think of this notion of 'aboutness' in terms of some constrained notion of causation or not, the fact remains that there is a clear sense in which Kenneth was, at one time, thinking about Fred, despite not being aware of it. This means that we can differentiate between what someone *thinks* their thoughts are about and what the thoughts are *actually* about. Nothing in my argument depends on theoretical matters as to how this distinction should be understood. Hence I will rely on our intuitive grasp of this distinction when claiming that Russell and Frege confused conventionally determined content with *beliefs* about conventionally determined content. In the sense above, they actually thought about one thing, despite taking themselves to be thinking about something else entirely. Alternatively phrased, they confused the beliefs that guide our use of a name with the convention governing the use of the name.

The appropriate method when trying to ascertain what an author is actually thinking about involves determining the most essential features of the author's view and then attempting to find the objects in reality that have, or could plausibly be thought to have, these features. I will follow this method and show that it gives a definitive result, namely that Russell and Frege were *actually* thinking about a subject's beliefs about the conventional reference of names.

3.2.2 The speaker's referent, conventional referent and supposed conventional referent

Consider the following three truisms that are rarely, if ever, denied. Firstly, there are public conventions that determine the meanings of words in general and also determine the referents of proper names. In this way it is a convention that 'Quine' refers to Quine and not to Ponting, just as it is a convention that 'ashes' denotes ashes and not shavings. This leads to the definition of the 'conventional referent'.

Conventional referent: The conventional referent of a name N is the individual that is determined in virtue of the communal convention governing the use of N.

The second truism is that language is *public*, i.e. communal, which implies that it has to be taught to individual speakers. The conventions governing the use of terms are empirical facts, in some sense, and individual speakers can learn these empirical facts correctly or incorrectly. In this way I can wrongly think that 'Quine' refers to Ponting, just as I can wrongly think that 'ashes' denotes shavings. This implies that individual speakers must have linguistic beliefs about the public meanings of terms and that these beliefs must include beliefs about the conventional referents of proper names. This leads to the definition of the 'supposed conventional referent'.

Supposed conventional referent: The supposed conventional referent of a name N for a speaker S is the individual that S believes to be the conventional referent of N.

The third truism is that proper names are used by speakers when they intend to draw attention to some object. This leads to the definition of the 'speaker's referent'.

Speaker's referent: The speaker's referent of a name N on an occasion of use and for a speaker S is the object that S intends to bring to the attention of his audience by using N.

Note that while the objects identified in these three ways upon an occasion of the use of a name will tend to be identical, they can diverge. Suppose say that Caroline and Drew are walking down the road when Caroline sees an individual, namely Siddle, behaving oddly. She wishes to warn Drew to stay away from the person behaving oddly. Looking closely, she gets a glimpse of the facial features of Siddle. Her senses are deceived, however, in that the *gestalt* of the facial features that her brief glimpse leaves her with is not that of Siddle, but of Quine. Matters are further complicated by the fact that, when she learned the name of the person with these facial features, i.e. the facial features that she thinks the person in front of her has, she learned it incorrectly. Instead of learning that the person with such features is called 'Quine', she learned that he is called 'Ponting'. Turning to Drew she now says: "Watch out, Ponting has gone crazy!"

In the above case the speaker's referent, the supposed conventional referent and the actual conventional referent are three distinct individuals. The speaker's referent of 'Ponting', i.e. the person Caroline was trying to bring to Drew's attention by uttering 'Ponting', is the man behaving oddly, namely Siddle. The conventional referent of 'Ponting', i.e. the referent de-

termined by the convention governing 'Ponting', is Ponting. The supposed conventional referent, i.e. the person that Caroline thinks that 'Ponting' refers to, is Quine.

The tripartite distinction constructed above gives rise to three distinguishable referential intentions. Caroline's main referential intention, in virtue of the conversational context, was to refer to Siddle, who is the speaker's referent of her utterance. The category of the supposed conventional referent gives rise to a secondary referential intention to refer to Quine. This referential intention is conceptually distinct from her primary referential intention to refer to Siddle. The category of the actual conventional referent gives rise to a third referential intention, due to the fact that Caroline presumably has a standing intention when using a name to refer to the conventional referent of the name[60].

The above example relies on a common-sense distinction between what people mean, what words mean and what individuals think words mean, as applied to names. With the distinction drawn, I can now state the exact claim that I wish to make. This is that traditional descriptivism is *actually* about the determination of the supposed conventional referent, i.e. about the relation between Caroline's utterance of 'Ponting' and Quine. Kripke's causal theory is about the determination of the actual conventional referent, i.e. about the relation between Caroline's utterance of 'Ponting' and Ponting. In other words, while the descriptivists were generally writing about the type of relation that obtains between 'Ponting' and Ponting, they confused this with the relation that obtains between 'Ponting' and Quine.

The traditional descriptivists confused supposed reference with conventional (semantic) reference, i.e. the phenomenon that led to the views of the traditional descriptivists is the fact that people have certain beliefs about what public language names refer to. The phenomenon that led to Kripke's views is the fact that words in public languages do have communally determined referents. The evidence that I will give for these claims is that the fundamental theoretical commitments and results of the descriptivist and Kripkean theories have exactly the properties that these phenomena have, or that they could plausibly be thought to have. My first argument concerns the result that would be yielded by an application of the theories of the traditional descriptivist and that of Kripke.

3.2.3 The argument from theoretical results (Dictionary argument)

---

[60] This point is based on two remarks made by Kripke (1977: 264, 273n). (The tripartite distinction developed above is my attempt at a more fine-grained distinction than the one between 'speaker's reference' and 'semantic reference' in Kripke's paper.)

Let us imagine that someone wanted to compile a dictionary of all the public names of persons that a given speaker - call her Emma - knows[61]. For every given name known to Emma our lexicographer includes a photo of the conventional referent of the name next to the name. Assume that the lexicographer uses the same general method to determine which name goes with which picture as lexicographers generally use when compiling traditional dictionaries. Call the completed result the *Conventional dictionary.*

Now imagine that the same lexicographer is also interested in the linguistic beliefs that Emma, at a specific point in time, has about specific names. Our lexicographer determines all the names that Emma uses, lists them, and, next to each, puts a picture of the individual that Emma believes the name to refer to. Assume that Emma has quite a few mistaken beliefs about the names that she uses. Call the completed result the *Emma dictionary.*

Now create two further documents by applying the traditional, descriptivist theory of names and the Kripkean, causalist theory of names to the names Emma can potentially use in utterances. In the case of Kripkean causalism, determine the individuals baptised at the beginning of the causal chains that led to Emma being in possession of the names she uses. Draw up a list of these names with a picture of the individual identified by the causal theory next to it. Call the result the *Kripkean dictionary.* Imagine that a semantic descriptivist performed the equivalent task for descriptivism. In other words, the descriptivist somehow identifies the descriptive conditions that, in terms of descriptivist semantics, Emma supposedly identifies with a given name, determines the person, if any, that these conditions apply to, and includes photos of the relevant person next to each name. Call this the *Russellian dictionary.*

The argument can now be stated very simply. The Russellian dictionary and the Kripkean dictionary will give differing results for a certain set of the names that Emma uses. Where these results differ *the Russellian dictionary will give the results in the Emma dictionary, while the Kripkean dictionary will give the results in the Conventional dictionary*. This is because the Conventional dictionary and the Emma dictionary will differ in cases where Emma uses a name incorrectly and precisely these facts about incorrect usage will (almost) always cause a similar divergence between the Kripkean and Russellian dictionaries. Simply put, where the person who compiles the Emma dictionary would explain divergence from the Conventional dictionary by saying that Emma has a false linguistic belief concerning a name, the person

---

[61] I am restricting this to persons, as countries are hard to take definitive pictures of and mathematical objects are even harder to take pictures of. Of course, nothing stops us from compiling non-pictorial dictionaries of such objects. In the case of mathematical objects, we already do this when we list the values of objects like Avogadro's constant, the Planck length, etc.

who compiles the Russellian dictionary would explain divergence from the Kripkean dictionary by saying that Emma attaches an idiosyncratic descriptive condition to a name. I think that the latter explanation is just a confused way of saying what is perfectly expressed in the former explanation. I will run through an example to illustrate such divergence.

Imagine the case of Emma who listens to a conversation between economics professors and acquires the names 'Paul Krugman' and 'Robert Lucas'. However, Emma did not understand the conversation well and managed to get the two mixed up. Emma acquires the belief that 'Lucas' refers to a person who won the 2008 Nobel Memorial prize in economics, authored *Peddling Prosperity* and is the most famous current defender of Keynes. In such a case the Conventional dictionary will still list Lucas as the referent of 'Lucas', but the Emma dictionary will list Krugman as the referent of 'Lucas'. This is quite straightforward and is just another way of saying that Emma incorrectly believes that 'Lucas' refers to Krugman.

The above difference will also be reflected in the Kripkean and Russellian dictionary. The Kripkean lexicographer will attempt to trace the use of 'Lucas' back to an original baptism. If we assume that the case of 'Lucas' is not deviant in some way, i.e. is not an Evans-style case, such a method would presumably identify Lucas as the referent of 'Lucas'. In a similar way, and even though there may be space for some disagreement as to exactly what should count as the relevant descriptive condition(s), the Russellian lexicographer will treat Emma's false linguistic beliefs as providing the descriptive condition she associates with 'Lucas'. The person that these conditions apply to is Krugman and for this reason the Russellian dictionary will list Krugman as the referent of 'Lucas'. In this way the Kripkean dictionary will (mostly) track the conventional dictionary, while the Russellian dictionary will track the Emma dictionary.

The individuals identified by the descriptivist theory of reference will correspond to the content of a dictionary drawn up to reflect Emma's linguistic beliefs. In the same way, the results of the causal theory will (mostly) reflect the results of a dictionary drawn up to reflect the actual conventional referents of the names that Emma uses. For this reason, I submit that traditional descriptivism is actually about (in my sense of 'about') our *beliefs* about the conventional reference of names, while Kripkean causalism is actually about the conventional reference of names. Alternatively phrased, the seemingly odd views of Russell and Frege can be attributed to the fact that, when considering names, they confused our beliefs about conventional reference with conventional reference itself. Call this the *dictionary* argument[62].

---

[62] It can, of course, be the case that some of our beliefs about what a name conventionally refer to will rarely, if

Four things should be noted about the above argument. Firstly, Kripke's theory will not give the exact results given in the Conventional dictionary. This is due to Evans's 'reference-switch' cases. Note, however, that all causalists, including Kripke (1981: 163), have responded to such cases by accepting that the causal theory is incomplete and that the data that there are such reference-shifts should be accepted. Their inquiries treat the explanation of the data in the Conventional dictionary as the ideal that they are supposed to strive towards when constructing theories. Hence the fact that Kripkean causalism won't quite generate the Conventional dictionary is no objection to the claim that Kripkean causalism is an attempt to explain the content of the Conventional dictionary.

Secondly, note that something similar holds with regard to the relation between descriptivism and the Emma dictionary. I find it plausible that, if a specific theory, say that of Russell, failed to match the Emma dictionary in certain ways, the descriptivist would view this as a problem for Russell's views. Any difficulty in identifying the individual that a person thinks a name refers to must also crop up as a problem for a descriptivist trying to determine the individual picked out in accordance with the descriptive condition governing a speaker's use of a name.

Thirdly, I wish to pre-empt any attacks based on the fact that Kripke's theory and traditional descriptivism are theories about the reference of names upon an occasion of use. This does not matter. We can imagine an extended version of the Conventional dictionary that captures the actual conventional referents of names as used by Emma over a period of time and an extended Emma dictionary that captures the supposed referents of names for the same speaker over the same period of time. The exact same reasoning would still apply.

Fourthly, the reader may well have noted that, if the argument is sound, it seems unlikely that we can find some topic on which Kripke and the traditional descriptivists both have reasonably plausible views. I will defend this contention later on, for I do not believe that there is a non-trivial question that traditional descriptivism and Kripkean causalism can be seen as remotely plausible rival answers to. We can reasonably inquire into how the conventional referent, the supposed referent and the speaker's referent are determined. These would be three conceptually distinct inquiries. There are also other concepts relevant to language that we can reasonably ask questions about. But there is simply no legitimate question that would generate

---

ever, be false. Consider, for instance, my belief that 'Ponting' conventionally refers to the person at the beginning of the causal chain from which I inherited 'Ponting'. Such beliefs will rarely be wrong, but such beliefs cannot, except in exceptional cases, exclusively determine supposed reference. For then it would render the truism that people are sometimes wrong about what a name refers to mysterious.

a dispute between Kripke and the traditional descriptivists in which neither is saying something absurd. If the term 'semantic referent' is thought to mean something distinct from the three concepts developed here (or obvious extensions of them), it is simply a will o' the wisp. I will proceed by looking at the theoretical commitments of causalism and descriptivism, in order to show that they match those we would expect from a theory of conventional reference and supposed conventional reference respectively.

3.2.4 The argument from eccentricity

The traditional descriptivists commit to *eccentricity*, i.e. the view that the descriptive conditions associated with a name by the utterer of the name determines the referent of that name. It has already been explained that this is an absurd view of proper names, which raises the question as to how it could have been seriously proposed. One potential explanation for the confusion would be to claim that the descriptivists confused speaker's reference with semantic reference, i.e. that they *actually* thought about the determination of the speaker's referent. The determination of the speaker's referent is uncontroversially eccentric, i.e. dependent only on what the utterer is trying to communicate. Furthermore, such referential intentions can plausibly be stated as descriptive conditions; we can say that someone intended to speaker-refer to 'the man in front of him', 'the richest man in the world', etc.

This, however, will not do. The conditions determining the speaker's referent when using a name may well eccentric, but it is too idiosyncratic. Such conditions differ for different speakers, which is consistent with what traditional descriptivists say about names, but they also differ wildly for one speaker from one occasion to the next, which is not. Of course, the traditional descriptivists, i.e. Russell, Frege and the developers of variants of their theories, are not quite in agreement on this issue. Both Russell and Frege are frequently portrayed as being committed to such stability in quick summaries of their work, but this does considerable violence to Russell's view (Russell, 1910: 115). Russell's actual view does not affect our argument though, and for this reason I will continue to speak of 'traditional descriptivism', so construed, for now. I will only discuss his actual view later on in this thesis.

The view of the traditional descriptivists (or, at least, the view commonly ascribed to them) is that there is a *standing meaning* that is associated with a name. This standing meaning will tend to stay relatively constant and only change upon occasion, presumably when very important information concerning the referent of a name comes to light. The conditions determining speaker's reference, however, can vary wildly from occasion to occasion, based on our communicative interests. Hence the content of traditional descriptivist theories of names

makes it highly implausible that descriptivist theories of references are, in my sense of 'about', about the speaker's reference.

We need to find some entity that can plausibly be construed as eccentric, but that is also relatively stable across occasions of use. The *beliefs that a given speaker has concerning the conventional referent*, i.e. the determinants of the supposed referent of a name in the speaker's vocabulary, fit the bill perfectly. I will demonstrate by using an example that illustrates my tripartite distinction, but that does not depend on perceptual beliefs as the Quine/Ponting/Siddle example did.

Consider a conversation where someone (Hans) says that, no matter how much money you have, you will always be unsatisfied as someone else will have more. It is, of course, not true that, for any given person, there is someone who is richer. Gerald wishes to communicate this by pointing out that the world's richest man would disagree with what Hans said. Gerald mistakenly believes that the chairman of Berkshire Hathaway is the richest man in the world. Furthermore, suppose that Gerald believes that the chairman of Berkshire Hathaway is called 'Gates'. The convention that 'Gates' refers to Bill Gates is a public convention that can be taught, and learned, incorrectly, and Gerald was unlucky enough to somehow get this wrong. Gerald now says: "I think Gates would disagree".

In order to fully understand the above situation we need to know that Gerald was trying to say something about the world's richest man - currently the telecommunications entrepreneur Carlos Slim - and hence the speaker's referent of his utterance was Slim. We also need to know that 'Gates' actually refers to Gates, i.e. that Gates is the referent conventionally assigned to 'Gates'. We also need to know that Gerald has a false belief concerning the convention governing 'Gates' (as well as the false substantive belief that the world's richest man is the chairman of Berkshire Hathaway) and that the person that he believes the name 'Gates' refers to is Buffett.

The notion of the speaker's referent captures the relation between Gerald's utterance of 'Gates' and Slim, while the notion of the conventional referent captures the relation between 'Gates' and Gates. In a similar way the notion of the supposed conventional referent captures the relation between Gerald's utterance of 'Gates' and Buffett. On the view defended, here Kripke's semantic theory is a theory of how it is that 'Gates' refers to Gates. Russell and Frege, however, without realising it, thought about the relation between 'Gates' and Buffett.

The notion of the supposed referent fits traditional descriptivism exactly. The determinants of the supposed conventional referents will differ across people as different people will have different, but mostly consistent, beliefs concerning the conventional referent of a name. This is due to the fact that different people will characterise their knowledge of the conventional referent of a name using different information. In this way, people who know someone from seeing them on a daily basis will probably use a visual stereotype to characterize their belief as to the referent of a name, people who have not seen a person will use other information concerning the person that they do have available[63], and famous people from long ago will typically have contemporary users characterize their beliefs concerning the referents of these people's names in terms of their 'famous deeds'. This is exactly the kind of thing that descriptivists claim for their *species* of meaning.

These determinants of the supposed referent will also be relatively stable across time, though not immutable. Over time our erroneous conventional beliefs can be corrected, we learn new and better criteria that identify certain individuals, and so on. This, once again, fits perfectly with what traditional, semantic descriptivism says about the conditions governing names. Furthermore, the determination of the supposed referent will be eccentric, i.e. depend on the individual, just as the referent determined in terms of traditional descriptivism is supposed to depend on the utterer only. Hence I think that supposed reference is what Russell and Frege were actually talking 'about', even though they did not realise it.

3.2.5 Interpreting Russell, again

In "Knowledge by Acquaintance and Knowledge by Description" (1910:115), Russell notes that different people can abbreviate the same name by using different descriptive conditions. He says that, for a single utterer trying to talk about a single individual, these descriptions will vary over time (114) and that it is a matter of chance 'which characteristics of a man's appearance will come into a friend's mind when he thinks of him' (114). Speaking of 'Bismarck', he says that 'the description in our minds will probably be some more or less vague mass of historical knowledge – far more, in most cases, than is required to identify him' (115). Later on Russell implies that all the information I have at my disposal does not apply on a given occasion of use. When writing of 'Julius Caesar' he says that, on a given occasion of use, 'in order

---

[63] In some cases the correct beliefs that two people have about what a name conventionally refers to may have nothing, or almost nothing, in common. If a blind and deaf person have sensory contact with Putin, then they can both acquire true, perceptually grounded beliefs about what 'Putin' conventionally refers to.

to discover what is actually on my mind when I judge about Julius Caesar, we must substitute for the proper name a description made up of *some of the things* I know about him' (119; my italics).

What are we to make of this? In terms of my analysis, these Russellian views can easily be interpreted as being *actually* about our linguistic beliefs. Such an interpretation would ascribe Russell's views to the fact that it can plausibly be claimed that there is never, or rarely, one consistently used descriptive condition that determines who I think a name refers to. Rather, what happens is that I wish to refer to someone, and, based on a piece of information already in my possession or subsequently acquired, I identify a name in my possession that I believe refers to this person. Such information then plays the role of 'linguistically relevant information' in a specific context.

Taking an example, suppose I spot someone across the street who is walking oddly. Looking closely I see the facial features of the person, realize that it is Harry who is walking oddly, and say 'Harry is limping'. In such a case the intended or speaker's referent is whoever is walking oddly. Note that I saw that the person walking oddly has certain facial features, and that this is what guided my use of the name 'Harry'. Hence it must be the case that I had the pre-existing, substantive belief that Harry has these facial features. This substantive belief guided my linguistic behaviour and hence the belief that whoever has those facial features is called 'Harry' can be ascribed to me. Hence, *at this given point in time*, 'Harry' is incontrovertibly tied to the condition 'person with such and such facial features'.

This only makes Harry's facial features the condition governing my use of 'Harry' in this case though. Next time I may wish to refer to 'the person I am hearing in the distance', recognize Harry's voice, and use my knowledge of what his voice sounds like to guide my use of 'Harry'. In other words my belief concerning *this* use of 'Harry' is best expressed as saying that I believe 'Harry' to refer to 'the person with such and such a voice'. In this way, any number of my substantive beliefs concerning Harry can play the role of determining my use of 'Harry' on a specific occasion of use. Sometimes, of course, the condition determining the intended referent will consist of a piece of information I already know to be true of some specific person and hence this information can, itself, serve as the condition governing linguistic use. In other words, I can intend to refer to my best friend and already know that Hogan is my best friend, and hence the content of 'Hogan' can be given by 'my best friend'. But in such a case the intended referent and the supposed referent are still *conceptually* distinct. These two aspects are always distinct in principle and sometimes distinct in practice. Hence the specifics

of Russell's view, i.e. the detail beyond his commitment to eccentricity, make perfect sense when interpreted as being due to a confusion on his part whereby he *actually* thought about our linguistic beliefs.

3.2.6 The argument from idiolects

The last argument I wish to make is quite short and concerns the matter of 'idiolects'. Traditional descriptivists are sometimes said to characterise language as a set of distinct, but overlapping 'idiolects'. I have already explained why we cannot get very far by interpreting talk of 'idiolects' as talk about personal linguistic conventions, even if this may be the standard usage of 'idiolect'. Given that this interpretation is off the table, what sense can we make of the claim that the traditional descriptivists were concerned with personal 'idiolects'?

Given that it is uncontroversial that linguistic conventions exist, and given that a speaker must have a set of beliefs concerning these conventions, I see no candidate for the idiolect of an individual other than identifying this idiolect with *the set of beliefs that a speaker has regarding these public, linguistic conventions*. When talking about 'idiolects' people typically do not seem to be talking about a subject's purely personal linguistic conventions. Such personal conventions are rare, and talk of 'idiolects' typically seem to assume that all people have such 'idiolects'. I think a more likely interpretation of what idiolect talk is actually about is that people who talk of idiolects are thinking of what we may term a speaker's 'linguistic grasp'. The most obvious way to characterize a speaker's 'linguistic grasp' is to equate such a grasp with the set of beliefs that the speaker has concerning conventional, public rules. In other words, I interpret idiolect talk as talk about private grasps of public rules. Hence, if traditional descriptivists were trying to characterise 'idiolectical meaning', and if idiolects are given by a subject's linguistic beliefs, then Frege[64] and Russell were actually talking about (in my sense of 'about'), our linguistic beliefs.

It has now been argued that both the commitment to eccentricity and the use of descriptive conditions that characterise traditional descriptivism can be neatly explained if we understand Russell and Frege as actually thinking about the determination of supposed reference. If the canonical interpretation of Russell and Frege is correct, so that their theories were supposed to be about semantic reference, then we must say that their theory of names confuses semantic

---

[64] Frege's claim that senses are public does not contradict his view that the users of names tend to speak idiosyncratic idiolects. The first claim amounts to saying that grasping a thought is a matter of being in touch with a public, abstract, objectively existing entity. The latter claim is the denial that people typically grasp the *same* public thought when uttering the same sentence containing a name.

reference and supposed reference in the same way as Kenneth confused Simon with Fred. This implies that there is a real sense in which Kripke and the traditional descriptivists were not thinking about the same thing. Hence we should not be surprised that the descriptivists view is about Kripke's topic, i.e. conventional reference, is absurd. Whereas the descriptivists were trying to give a view about conventionally determined reference, they confused this with a subject's beliefs about conventionally determined reference and so ended up proposing a very odd doctrine about conventional reference.

3.2.7 An objection

One way to object to my view would be to claim that I have misconstrued the dialectic, and to claim that there exists some question such that neither Kripke's view, nor the descriptivist view, is an absurd answer to it. On such a view, then, I have erred in equating 'semantic' reference with 'conventional' reference; rather 'semantic' reference is something else entirely. Such a dialectical move is misguided, as there is no sense of 'semantic' in which we do not have to convict either the descriptivists or Kripke of making an absurd claim.

A seemingly plausible objection, however, can go as follows. Let us define the 'semantic referent' of a name as the individual that best explains our attribution of truth-conditions to an utterance in which the name occurs. This leads to a well-defined research program: determine what truth-conditions competent speakers attribute to utterances, and then construct a theory that explains such attributions. Using this definition, it is perfectly possible that we can discover that interpreters assign truth-conditions in one of at least three different ways.

We can discover - the objection continues - that interpreters assign truth-conditions to utterances in which names occur based on the causal chain of use relevant to the name, or based on the descriptive condition the utterer associates with the name, or, for that matter, based on what the utterer was trying to refer to when using the name. If we discovered such a thing it could show that, while people do use public conventions in order to communicate, these conventions *do not determine the semantic referent*, as defined above. Call a language where interpreters assign truth-conditions based on the descriptive condition the utterer attaches to the name *Russellian English*, call a language where interpreters assign truth-conditions based on the causal chain from which the utterer inherited the name *Kripkean English*, and call a language where truth-conditions are assigned in terms of what the speaker - based on the conversational context – is trying to refer to *Donnellian English*. On this definition of 'semantic reference' it is clear that Russellian English, Kripkean English and Donnellian English are rival answers to an empirical question. We can give subjects cases like Ponting/Quine/Siddle or

Slim/Gates/Buffett and try to determine whether English is Kripkean, Russellian or Donnellian.

3.2.8 Objection answered: the objector's definition of 'semantic' cannot be used to ask a determinate question

I think the above objection captures how a majority of philosophers think of 'semantic content'. It is assumed that interpreters assign truth-conditions to utterances and that we should try to develop theories that describe and explain these judgments. The problem with such a view lies in the very first stipulation made by the objector, namely '[l]et us define the 'semantic referent' of a term as the individual that best explains our attribution of truth-conditions to an utterance'. This does not amount to a well-formed definition with a determinate content of the required type. An interpreter can take us to be asking about the truth-conditions determined by what the speaker means, what the speaker believes the words mean, what the words actually mean, or some other related type of 'meaning'. *But we already know the answer to the first three questions*, and the definition itself fails to identify a fourth question. We already know what the conventional, supposed and intended referents are in cases like Gates/Buffet/Slim and Ponting/Quine/Siddle. We also know, as I have argued, that Russellian English can no more be a plausible theory of what words mean than Kripkean English can be a plausible theory of what speakers mean or what people think words mean. However, if we force the interpreter to answer without explicitly telling him what we are asking about, all we will learn from his answer is *how he interpreted our question*, and not any deep truths about 'semantics'. This is analogous to the way in which anyone who assigns Donnellan's (1966) truth-conditions to a misdescription-case is only revealing that he took us to be asking about speaker meaning. Such data reveals linguistic ambiguity, not philosophical depth.

An analogy can serve to drive the point home. Let's say I wish to determine how the word-form 'meter' is to be pronounced. Suppose that I have heard people pronouncing the word in radically different ways, and wish to discover which one was standard, i.e. I wish to discover how the vast majority of the population pronounces it. This sounds like a well-defined research-project. I can ask people to pronounce the word, record the results and determine which pronunciation is most common.

Let us suppose, however, that I had the misfortune of only asking trilingual speakers who are fluent in Dutch, English and German. The problem, then, is that the word-form 'meter', exists in English, Dutch and German, and means the same, but is pronounced differently. My results will be completely useless. If the English pronunciation 'wins' this will not reflect the fact

that it is standard. Rather it will reflect the fact that *the majority of the test-subjects took me to be asking about English pronunciation.* My results will not establish the 'proper pronunciation'. Indeed, it could not do so, as there is no such thing as the 'standard pronunciation of "meter"', unless relativized to a language.

The question 'How do you pronounce this word?' can be interpreted as 'How do you pronounce this English word?', 'How do you pronounce this Dutch word?' or 'How do you pronounce this German word?'. These are three independent questions with independent answers. If 'How do you pronounce this word?', as used here, is supposed to mean anything distinct from this it is simply meaningless, but deceptive, babble. Herein lies the analogy (and I think it is an exact analogy to what has actually happened in semantics). The ready availability of answers to the question 'What are the truth-conditions of this utterance?' only indicates that people take such a question to concern one of the notions discussed. But, just as there is no 'real' question as to the standard pronunciation of 'meter', so there is no question as to the 'real' truth-conditions. The question can, of course, be used to discover which one of the kinds of meaning someone took us to have meant by 'meaning' when we asked this question. But this trivial question is plainly not what has driven more than a century of semantic inquiry.

3.2.8 Other definitions of 'semantic' that suffer from essentially the same defect

In the objection above the objector tried to define the term 'semantic referent' with reference to the truth-conditions that competent speakers would attribute to utterances. The problem was that the 'truth-conditions' need to be truth conditions *of something*, i.e. truth-conditions of the conventional meaning of the utterance, truth-conditions of the supposed linguistic meaning of the utterance, truth-conditions of the intended meaning of the utterance, or truth-conditions of some related type of 'meaning'. The question itself, as asked by the objector, does not succeed in determining a *species* of meaning for us to have disputes about. There are a myriad of ways in which an objector can formulate his definition of 'semantic' that would run afoul of the exact same requirement. I cannot go through all superficially different attempts that I have encountered that suffer exactly the same defect, but will merely mention four.

The real issue between traditional descriptivism and causalism is about how sentences manage to represent their intentional objects. Here we have the same problem again. The objector needs to specify whether he is talking about how sentences represent their conventionally assigned intentional objects, how sentences manage to communicate the intentional object that

the speaker has in mind, how the sentence is related to the intentional object that the speaker believes the sentence to represent, or some related question. On the first three readings traditional descriptivism and externalist causalism cannot both be (even remotely) plausible rival theories, and the formulation itself does not succeed in identifying a fourth option.

The real issue between traditional descriptivism and causalism is about the appropriate input for a compositional semantic theory that attempts to explain communication. Again, the question to be asked is 'compositional theory *of what*?' We can have a theory of the conventional meaning of a sentence, the supposed meaning of a sentence and the intended meaning of a sentence, and all of these are essential to explaining communication. On each of these readings traditional descriptivism and externalist causalism cannot be rival theories worth taking seriously, and the formulation itself does not identify any fourth type of 'content'[65].

Theories of semantic content are about the propositional content ascribed in that-clauses. Such content serves as the objects of beliefs and other attitudes. Causalism and traditional descriptivism are primarily disputes about such contents. Here we still have the same problem. The tripartite distinction developed earlier applies to expression containing that-clauses, as for any expression that includes a that-clause of the appropriate type we can still distinguish between what the that-clause conventionally ascribes, what the speaker thinks the that-clause in question conventionally ascribes and what the speaker was primarily trying to ascribe by using the that-clause. Here all three of these entities are ill-understood. But, be that as it may, the proposed definition lacks a clear sense until disambiguated in this manner. And, if disambiguated, it is again plain that traditional descriptivism and causalism cannot be serious rival theories of any of these three *species* of content, and the formulation itself fails to identify a fourth kind of 'content'.

The real issue between traditional descriptivism and causalism is about what should or should not be values of names in a formal semantic theory. We can have a formal theory of the conventional meaning of sentences, a formal theory of the supposed (idiolectical) meaning of sentences, a formal theory of the primary (intended) meaning of sentences and a formal theory of other, related notions. We can then, concerning each of these three topics, have a view concerning reference, i.e. a view of how the names in question acquire whatever their values may be. Once again, this definition of 'semantic' is of no value until it has been spe-

---

[65] Strictly speaking, of course, such an objection concerns semantics, whereas this thesis concerns reference. But, even if we rephrased the objection so that it concerns the question of how names in such a theory acquire their reference, the problem remains.

cified what the formal theory in question is a formal theory about, whether it is one of the objects of my tripartite distinction or something else. On the first three readings, traditional descriptivism and externalist causalism cannot be plausible rival theories of reference and no fourth option has been identified.

In all of the above cases the problem is that the objector's definition of 'semantic' (or 'intentional object', etc.) underdetermines what the topic of inquiry is supposed to be. Such objections suffer the same defect as the one discussed at length and are irrelevant. Also note that a need for a theory of the objects of my tripartite distinction, i.e. a theory of what words mean, what people mean and what people think words mean, arises naturally from any inquiry into communication and that what such inquiries are about could easily be explained to a ten-year old. This is in sharp contrast to the above, putative definitions of 'semantic'. Their seeming sophistication only serves to conceal a lack of conceptual depth, rigour and clarity. No doubt a hundred different semanticists can provide a hundred slightly different definitions of 'semantic', but I fail to see one that does not run into exactly the same difficulties.

Traditional descriptivism can no more be a *plausible* theory of conventional reference than Kripke's views can give a plausible account of the intended or supposed referent. This means that we can only save the idea that there is a clash between their views that is worth taking seriously if we find another worthwhile topic concerning names for them to disagree on. The way philosophers typically define 'semantic' falls a long way short of identifying such a topic and I fail to see a candidate that would fare any better[66].

3.3 Why did Russell and Frege confuse semantic reference and supposed reference?

The traditional descriptivists, *contra* Sainsbury's doubts about Russell, took themselves to be presenting a theory of conventional reference. The resulting theory, however, is absurd. I have argued that such absurdity is probably due to Russell and Frege's confusing the determination of conventional reference with the determination of supposed reference. This claim, however,

---

[66] My own suspicion as to how this matter went unnoticed is that it is due to the several *species* of 'meaning' that we have an intuitive grasp of and that are relevant to explaining communication. When we read Russell and Frege, we automatically interpret their claims as being about a speaker's linguistic beliefs, without recognizing it in these terms. In other words, we suffer the same confusion that I am accusing them of. When we read Kripke we, based on what he explicitly says and the truth-values he ascribes to his example-cases, similarly immediately interpret his claims as being about conventional reference, again without recognizing this in these terms. When we *then* look back at traditional descriptivism, interpreting it as being about conventionally determined reference, it becomes obvious that it is badly wrong.

even if true, does not fully scratch our explanatory itch. If, indeed, such confusion explains their doctrines, then how did this muddling of distinct topics occur? Here, my view will inevitably be somewhat speculative. Yet I think there is enough evidence for the view that it is worth proposing. I think that the basic origin of the trouble is in their assumption that language generally exhibit cognitive-semantic identity, i.e. in the view that if a competent speaker uses a sentence, then the thought in the mind of the speaker will have the same content as the content of the sentence. If we apply this view to a sub-sentential entity like a name, and assume compositionality, it implies the view that *part of the thought in the mind of an utterer of a name will have the same content as the content of the name*. I have already argued that both Russell and Frege can reasonably interpreted as committing, if implicitly, to such a doctrine. How, then, does this doctrine explain their semantic claims about names?

Note that there are a lot of sub-sentential expressions where it is plausible to claim that part of the thought in the mind of the utterer in has the same content as the expression itself. The conventional content of 'university' can plausibly be claimed to be 'tertiary educational institution that awards degrees' and we can also plausibly claim that if a competent speaker uses the term 'university', then they have some thought, the content of which includes 'tertiary educational institution that awards degrees'. The same goes for 'chair' and 'piece of furniture designed to be sat on', 'MVP' and 'most valuable player', and so on. Such terms can reasonably be thought to exhibit cognitive-semantic identity.

It became increasingly apparent, as 20th century semantics progressed, that all terms do not exhibit cognitive-semantic identity. The matter of cognitive-semantic identity is somewhat complicated in the case of indexicals, where, even though a version of the claims can be maintained, it is apparent that the cognitive content, conventional content and propositional contribution of indexicals must be distinguished. Cognitive–conventional identity breaks down even more dramatically in the case of conventions defined in terms of real world objects. Stipulate that there is a person, and we decide to conventionally speaker-refer to him by using the term 'Glob'. The content of our convention can only, using our language, be expressed as 'Use "Glob" to speaker-refer to Glob'. Yet, if I showed this expression of the convention to someone, they will learn nothing useful above and beyond the knowledge that 'Glob' can be used to conventionally refer; it is only once they have some way of identifying Glob that they can usefully employ 'Glob'. Speakers will differ in their ways of identifying Glob. One speaker may associate some visual stereotype with 'Glob', another may identify Glob by what he sounds like, another may know that Glob is the tallest man in England, and so on. Hence these speakers will all follow the same rule, namely 'Use "Glob" to speaker-

refer to Glob', yet they will all do so in virtue of following distinct, and defeasible 'proxy-rules'. Rules like 'Use "Glob" to speaker-refer to the person who matches visual stereotype *a*', 'Use "Glob" to speaker-refer to the tallest man in England', and so on, will allow users of 'Glob' to participate in the convention governing 'Glob'. These proxy-rules, however, do not give the content of *several unique* conventions. Rather they are just defeasible strategies used to follow the *single* convention 'Use "Glob" to speaker-refer to Glob'[67].

Note that dictionaries typically give us the conventional content of a term, but only if this is useful. When they do contain information about names, as they occasionally do, they do not give the useless '"London" refers to London', but the rather more useful, salient proxy-rule '"London" is the capital of England'. The same goes for their treatment of natural kind terms. This, unfortunately, muddles the distinction between conventional content and commonly used proxy-rules. In doing so dictionaries make it easier to suppose that all terms exhibit cognitive-semantic identity.

The cognitive content of some term and the conventional content of it will come apart whenever the conventional content is defined in terms of some real world object that can be identified in several ways. Call any such convention *object-dependent* conventions. Note that Kripke was able to attack the descriptivists on the matter of names and natural kind terms, i.e. terms where it is very plausible to claim that our actual conventions are object-dependent. Also note that there is nothing special about our linguistic practices that allow cognitive and conventional content to come apart. Consider an iterated coordination game in which subjects have to kick the same ball, or write on the same wall, or shout at the same person, etc. In all such cases we can expect an object-dependent convention, defined in terms of the specific ball, wall or person, to emerge. We can, similarly, expect the subjects to follow different, defeasible proxy-rules in order to follow a single convention with a single content, i.e. proxy-rules like 'Kick the ball in the corner', 'Kick the red ball', 'Kick the ball that satisfies visual stereotype *a*' and so on. Hence we are here dealing with a general phenomenon that rears its head in language, not some language-specific matter.

Cognitive-semantic identity, as construed above, will *always* fail in the case of object-dependent linguistic conventions. We could, of course, avoid such failure. Cognitive-semantic identity can be turned into a truism by allowing the cognitive content to be individuated broadly.

---

[67] These proxy-rules must, of course, be rigidified in some or other way. Nothing here depends on how this is to be done, so I will ignore this issue here.

By such a standard, it could be allowed that two people can have the same singular thought, even if they follow different proxy-rules and their thoughts have different narrow contents. This, however, does not affect my argument against Russell and Frege. The examples of descriptive conditions given by Russell and Frege make it plain that they did not adopt such a broad standard for individuating thoughts[68]. Strictly speaking, however, this implies that we should not say that the trouble was merely their commitment to cognitive-semantic identity. Rather the problem was their commitment to cognitive-semantic identity, *given* their narrow standard for the individuation of thought.

My speculative diagnosis, then, is as follows. Despite using a narrow standard for the individuation of thought, Russell and Frege implicitly assumed that all terms exhibit cognitive-semantic identity. If one adopts such a standard for the individuation of thought, and assumes cognitive-semantic identity, then Millianism about the semantics of names is immediately off the table. When they turned their attention to names, Russell and Frege noticed that, even in the case of competent name-users, such users typically associate eccentric, i.e. individually variable, *cognitive* contents (the proxy-rules which guide usage) with a specific name. They then, based on a prior commitment to cognitive-semantic identity, wrongly took the content of these proxy-rules to also be *semantic* contents. Given that such proxy-rules can differ even when the same convention is followed, their equation of thought content and semantic content forced them to say that names have eccentric semantic content. In this way they mistook the determination of supposed reference for the determination of conventional reference.

4. Conclusion

In this chapter I have claimed that traditional descriptivism is false as a theory of semantic reference. This can be seen from the difficulty it has in dealing with cases where utterers associate the wrong descriptive conditions with a name, and should ultimately be blamed on traditional descriptivism's commitment to eccentricity. The commitment to eccentricity is absurd enough that we need to explain how it came about. I have argued the data that descriptivist theories try to account for matches the data that an inquiry into the linguistic beliefs of a speaker would generate (the Dictionary argument) and that the theoretical commitments of traditional descriptivism matches plausible views about the linguistic beliefs of a speaker (the Eccentricity argument). From this it follows that Russell and Frege's views must stem from

---

[68] Russell individuates thoughts strictly in terms of objects that the subject is acquainted with (1910: 117). Kripke (2008) has claimed that Frege is also committed to something akin to Russellian acquaintance. If this is right it may help to explain why Frege adopted a narrow standard for individuating thoughts.

thinking about what we sometimes call the 'idiolects' of individual speakers. Such idiolects must (overwhelmingly) consist of a speaker's beliefs about the public conventions governing terms in a public language, and hence this is what they, in some sense, must have confused with semantic reference. I blame such confusion on their commitment to the idea that, in the case of competent speakers of a language, the (narrow) content of the thought that guides their utterance of a sentence will be identical to the semantic content of the sentence itself. This idea, while not implausible in where certain sub-sentential expressions are concerned, is false; thought content and semantic content will come apart in the case of object-dependent conventions. Speakers who follow object-dependent conventions, however, have a need for proxy-rules in order to do so. These proxy-rules have a form that is misleadingly similar to the content of conventions giving the use of terms that do exhibit cognitive-semantic identity. This makes it easy to mistake different proxy-rules used by different speakers to follow a single convention for a set of distinct conventions. In this way, then, Russell, Frege and the later descriptivists were led down the garden path.

It has now been argued that the appeal of both causalism and descriptivism rests on conceptual confusions that can be cleared up by considering the nature of conventions. In the next chapter I discuss matters that arose at various parts of the discussion thus far, but could not be profitably discussed without taking us too far afield from more immediate concerns.

Chapter 4: Some loose ends and a conclusion

1. Introduction

In this chapter I will discuss two issues that arise from the preceding discussion. The first concerns the fact that theories like the coordination view get little discussion in the philosophical literature and the second issue concerns the usefulness of thinking about semantics in game theoretical terms. I will end the chapter by explaining what I take to be the main lessons to be learned from the discussion in this thesis.

2. The strange case of the missing theory of reference

The coordination view of the semantic reference of names is, as far as I am aware, a novel view of semantic reference. It is, of course, just one way of trying to define semantic reference in terms of speaker's reference and some notion of convention. If the Lewisian is unmoved my by arguments for the coordinating rule view of conventions, he could simply plug in a Lewisian view of conventions and arrive at a different foundational view of semantic reference. This would be another variant of what we may call *Speaker's referent – Convention* views of semantic reference (hereafter SRC views).

> SRC view: A name *N* refers to a particular *o* in a linguistic community *L* if, and only if, *N* is conventionally used among *L* to speaker-refer to *o*.

We could generate different SRC views of the reference of names with different definitions of convention or speaker's reference. Such SRC views are, of course, quite similar in structure to conventional theories of sentence meaning. These theories typically define sentence meaning in terms of speaker meaning and convention, to yield some variant of the following basic idea, which I will call the *Speaker Meaning – Convention* view (or SMC view):

> 'SMC view': A sentence *s* means that *p* in a linguistic community *L* if, and only if, *s* is conventionally used among *L* to speaker-mean that *p*.

SMC views are typically developed by combining a Gricean view of speaker-meaning (Grice, 1957) with a Lewisian analysis of conventions. It is not difficult to understand the popularity of SMC theories of sentence meaning. It rests on two basic ideas, namely that language is conventional and that such conventions enable speakers to communicate what they mean. These ideas are simple and accords with common sense. While it has proved difficult to flesh

out these basic notions, the simple and intuitive nature of SMC theories gives it a perennial appeal. This raises a puzzle, namely why one struggles to find SRC theories of reference in the literature[69].

The single most interesting property of SRC views of semantic reference at this point would seem to be the fact that they do not feature in discussion concerning the semantic reference of names. Both SRC and SMC theories use the notion of a convention in the same way, and hence it cannot be some difficulty with conventions that explains the absence of SRC views. The only difference between the two lies in the fact that SMC views rely on a notion of the speaker meaning of a sentence, whereas SRC views rely on the notion of the speaker's referent of a proper name. This, however, cannot account for the absence of SRC views, as the notion of the speaker's reference of a proper name is not obscure or problematic in some way that the notion of the speaker meaning of a sentence is not. In fact, the notion of speaker's reference is, if anything, more intuitive than the notion of the speaker meaning of a sentence. What's more, the most highly regarded theorist about semantic reference, namely Kripke himself, has stated that the distinction between speaker's reference and semantic reference is applicable to all languages (1977: 267).

The basic building blocks of an SRC view, namely speaker's reference and convention, are generally thought to be legitimate among philosophers. What's more, it is intuitively plausible to state that speaker's reference relates to semantic reference in the same way that the speaker meaning of a sentence relates to the semantic meaning of a sentence. It may be objected that the SRC view is about reference, while SMC views are about content. But this makes little difference, as we could, and probably should, view the semantic content of a name as its referent. Hence the SRC view of semantic reference leads directly to an SCR view of the semantic content of a name.

The absence of SRC theories of semantic reference becomes even stranger if we look at the possible theories that are generally discussed when we look at encyclopaedia entries on the two subjects. Here we typically find that considerations from the theory of semantic reference

---

[69] Davis (2005) has proposed roughly a Lewis-Grice theory of names. Davis, however, is a Lockean who holds that language expresses ideas. He states that the content of a name is an 'atomic idea' that is expressed by a name, and that "the word 'Aristotle' is meaningful because it is conventionally used to express the idea of Aristotle" (2005: 8). As should be clear from the views I have defended, however, I see no reason to think that ordinary proper names express ideas or to think that we need to posit such ideas to explain how we use names like 'Obama', 'Tendulkar', etc.

have found their way into discussions about sentence meaning, but that such influence has been mostly one-way traffic. The reverse, i.e. attempts at constructing intention-convention theories of semantic reference, does not feature prominently. Compare, for instance, the *Stanford Encyclopedia of Philosophy* entries on "Reference" (Reimer, 2010) with the entry on "Theories of Meaning" (Speaks, 2011). The article on semantic meaning contains a section on attempts to develop a causal theory of semantic meaning that are inspired by Kripke's causal theory of semantic reference. The reverse is not the case as no mention is made of Grice, or Lewis' *Convention*, when matters concerning semantic reference are discussed.

We may think that the difference between sentences and subsentential entities like names explains the above discrepancy. But there is nothing particular to communicative intentions, or conventions, that restrict their explanatory relevance to the meaning of sentences. The relevance of Gricean views to understanding the meaning of sentences derives from the intuitive idea that speakers typically utter a sentence with the intention to bring some proposition that the speaker has in mind to the attention of his audience. It is similarly intuitive to claim that speakers typically utter a referential, subsentential element, for instance a name, with the intention of bringing some individual that the speaker has in mind to the attention of his audience. The relevance of a broadly Lewisian construal of conventions to the understanding of sentence meaning derives from the platitude that the meaning of a sentence is determined by convention. It is similarly a platitude that the referent of a name is determined by convention. Despite this there is no commonly discussed intention-convention view of reference. Note that, in fact, the nature of SRC views allows them to avoid the most common complaint against SMC views, namely that the existence of an infinite number of sentences commits the SMC theorist to the existence of an infinite amount of conventions. Given that SRC views concern sub-sentential entities like sentences, as opposed to sentences as such, this concern does not affect SRC views[70].

It seems clear that the absence of such theories must be due to historical and sociological factors. When we look at the history of how the two subjects developed, we find that the history of theorising about semantic reference looks remarkably different from the history of theorising about sentence meaning. The history of theorising about semantic reference has mainly taken the form of a dispute between various descriptivist views (that take their basic

---

[70] SRC views may, of course, face the converse problem of accounting for 'the unity of the proposition'. If, indeed, this is a legitimate concern.

inspiration from Russell and/or Frege) and various causal views that derive from Kripke's *Naming and Necessity*. Given the form that such theories take, it is not obvious how such views are supposed to relate to issues concerning speaker meaning and, more importantly, conventions[71]. Furthermore, theorising about semantic reference has also generally been driven by concerns about how to formalise the logical structure of natural language, whereas theorising about sentence meaning has been driven by the attempt to understand communication and, to some degree, understanding. These projects have mostly been carried out independently from one another with no attempt being made to determine whether the popular views offered on these topics are consistent. In fact, the degree to which these issues have been studied independently of one another is remarkable. In many ways it may as well have been done by groups that do not know about the existence of the other. Even where the same philosopher has addressed both questions, such thinkers have seemingly treated them as unrelated concerns. A paradigmatic case is Lewis' attempt to defend a causal version of descriptivism (1984: 226 - 229). Lewis puts forward causal descriptivism without making any reference to his work on sentence meaning - or conventions in general – despite the *prima facie* plausibility of simply adapting his SMC view about semantic meaning so that it becomes an SRC view about the semantic reference of proper names.

If the main claims in this thesis are correct, then those who theorised about communication have understood their subject better than those who were motivated by concerns originating in logic. The fundamental constraint in theorising about communication is that one has to account for how information gets from one place to another. This question amounts to a well-formed question, i.e., apart from various issues about how the notion of 'information' should be conceptualised, it is a question with a definite sense. In fact, this is the question that, in its most abstract form, inspired the revolutionary work done by Clause Shannon (1948) and others. If my claims in chapter three are correct, however, then those who theorised about reference have suffered in virtue of the fact that the question they were trying to answer was not clearly understood. The question is typically formulated as a matter of trying to give a semantic theory that explains the truth-conditions that competent users assign to utterances. But, as already argued in chapter three, such a way of thinking about semantics does not amount to a well-formed question. Differing intuitions about such cases generally reflect intuitions about

---

[71] In my experience causalists think that the fact that causalism is consistent with the idea that some baptised particular could have been baptised by using a different name is enough to make causalism consistent with any plausible theory of convention and hence that there is nothing to be learned from considering the nature of conventions as such. This view, if widespread, is probably what led to them to mistake trivialities about conventions for revelations about reference.

different things. My speculative diagnosis is that this is the core fact that accounts for the fact that SRC theories are generally missing from the literature concerning semantic reference. It has resulted in a situation where one of the two most prominent theories of semantic reference is not actually a theory of reference at all, while the other is trivial.

3. 'Gödel'/'Schmidt', equilibrium selection rules and power-users

3.1 Inequality in the linguistic democracy

In this thesis I have put forward the coordination view of semantic reference, according to which a name semantically refer to the particular that, absent false beliefs and defeaters, all members of a linguistic community are disposed to use the name to speaker-refer to in virtue of a coordinating rule that advises them to do so. The first challenge to such a view, of course, is that it stands in need of a theory of mental content, with the most basic constraint being that such a notion of mental content cannot make an irreducible appeal to any notion of conventional reference. This problem is a familiar one, of course, and a standard issue in the debate concerning the idea of a language of thought. I have no opinion on that debate that would fit well into the present work.

The second challenge facing the coordination view concerns a statement of the conditions that determine the adoption of semantic rules. If the coordination view of semantic reference is correct, then this amounts to a game theoretical problem in the way indicated previously: language users can be thought of as involved in a meta-coordination game in which they are trying to adopt the same rules. Such a situation requires a procedure, or criterion, for choosing between an infinite amount of possible equilibria, i.e. the situation requires what game theorists term an 'equilibrium selection rule'. The situation is complicated somewhat in that, as indicated previously, the typical case where someone has to decide whether to adopt a rule or not it has some of the features of an asynchronous game as those who already use the rule have a *prima facie* interest in keeping the rules that they already use. Furthermore, all users can change the rule that they follow, so that that it amounts to a continuous, as opposed to discrete move, form of such an asynchronous game.

One way to discern the equilibrium selection rule that people do use would be empirical study of actual language users. It will, however, be quite difficult to so. The best thing to study would be reference-switches and cases where there is some pressure towards reference-

switching in virtue of people using the wrong name to speaker-refer for an extended period of time. The problem with doing this is that such cases are quite rare. The other thing to look at would be cases where a baptism does not succeed and trying to see why it failed. The problem is that such cases often has to do with extra-communicative factors; attempts to give nicknames may fail because they do not seem 'fitting', people may ignore a suggested name change of a city for political or aesthetic reasons, and so on.

A different way of looking at the problem would be to retreat from empirical study, make various simplifying assumptions and consider what equilibrium selection rules would be optimal under such conditions. This should at least give an idea of the issues involved and the 'lay of the land', so to speak. I will briefly run through an example to show what I mean.

Consider an exaggerated and stylised version of Kripke's 'Gödel'/'Schmidt' problem. Stipulate that a linguistic community includes a hundred people, that they are fully rational and only choose linguistic rules based on communicative efficiency. Stipulate that seventy of them have never heard of Gödel or Schmidt and that the relevant facts are as Kripke indicated. Further stipulate that, among the thirty who have heard of Gödel, twenty have only heard of him as the author of his incompleteness theorems and also have no interest whatsoever in learning anything else about Gödel. When they use 'Gödel', they *always* intend to speaker-refer to the author of the two theorems. Further stipulate that all thirty people who have heard of Schmidt are suddenly told that it was proven by someone named 'Schmidt' and that all these facts are common knowledge between the thirty people. They now have to choose which rule to adopt as they have reason to believe that, at some unknown point or points in the future, they will wish to speaker-refer to Gödel.

In the above case I have stacked the deck so that the majority of twenty people should have a *prima facie* interest in sticking with their usage and effectively treating 'author of the incompleteness-theorems' as a reference-fixing description. We may be tempted to think that in such a case using 'Gödel' to conventionally refer to the person who actually proved the theorems would be optimal. Remember that, on the coordination view, such choices are pragmatic, i.e. what matters primarily is that all adopt the same coordinating rule. They could express everything they would wish to say by adopting either convention, the only question is which convention is best. It take it that it is common cause that, if we keep making the hypothetical scenario more extreme, the majority larger, the original mistake more distant in the past, and so on, eventually switching would be the intuitively best option. This, presumably,

is the lesson to be learned from Evans's mythical 'Ibn Khan', the mathematician who is conventionally referred to by the name of the scribe who recorded his work[72].

We may be tempted to think that an equilibrium selection rule like 'pick the option that the majority currently uses' is the best long run policy in order to maximize coordination. The answer, however, is not that simple. If we assume that everyone has a vested interest in sticking with their current use it follows that, in the linguistic democracy, all voters are not equal. Most names are such that a lot of people will utter them a few times and a small number of people will use them extremely often. Friends, family and loved ones will use one's name very often; a much larger number of people will use it a few times. Call users who use a name often 'power-users'. These power users, apart from being likely to have a much stronger interest in sticking with their usage, are also much more likely to be the people from whom people who have never heard a name acquire it. As such their usage is much more influential and likely to spread among the seventy people who were stipulated to have never heard of Gödel[73]. If we stipulate that the ten people in the stylized 'Gödel'/'Schmidt' example are power-users, i.e. mathematicians, biographers of Gödel and the like, it would be most rational to ignore the current majority and, on the assumption that the power-user use will probably be most prominent by the time one needs to speaker-refer to Gödel in future, adopt the rule of using 'Gödel' to speaker-refer to Gödel, and not Schmidt, as the majority have been doing.

The above case may be extremely artificial, but it allows us to see something important. If we assume that communication is a coordination game and that some people use a name a lot more often than others, this immediately leads to the conclusion that all votes are not equal in the linguistic democracy. Language users can be expected to defer to those who use a name often. Also note that this may give a partial explanation of why, in some cases, we defer to experts where natural kind terms are concerned, as experts will typically also be power-users[74]. More importantly though, I hope that the example demonstrates that thinking about

---

[72]An urn is discovered in which are found fascinating mathematical proofs. Inscribed at the bottom is the name 'Ibn Khan' which is quite naturally taken to be the name of the constructor of the proofs. Consequently it passes into common usage amongst mathematicians concerned with that branch of mathematics. 'Khan conjectured here that…' and the like. However suppose the name was the name of the scribe who had transcribed the proofs much later; a small '*id scripsit*' had been obliterated (Evans, 1982: 306).

[73] Note that everyone is a power user of their own name, and, except in the case of famous people, by far the most likely person to introduce new people to it. For this reason it would make sense to defer to people when they tell us their names.

[74] This is not supposed to be a rival explanation to Putnam's explanation of deference to experts (1975), but a supplementary one. Expertise as such will presumably also give one disproportionate power in the linguistic democracy. For a start, an expert's use of a medical term like 'arthritis' is more likely to be medically useful than

what would be optimal to do for rational beings in meta-coordination games can be of some use.

3.2 Assignment theories and equilibrium selection rules

In chapter two I argued that the Kripkean causal theory has no non-trivial explanatory content. Note that, on the above view, we actually *can* assign the causal theory some explanatory use. We could interpret the claim that a name conventionally refers to the particular baptised at the beginning of the causal chain as the formulation of an equilibrium selection rule. In other words, if speakers in a community do all they can to make sure that their use of a name is consistent with a causally related baptism, i.e. adopting only usage they think they have acquired in this way, abandoning usage that did not originate in this way and the like, then their practice will generate exactly the data predicted by the causal theory. In such a case the Kripkean causal theory, while not a foundational theory of reference in the way that reductive causalism is, would still be much more than the statement of a mere correlation. It would be genuinely explanatory as linguistic practice in such a community consists in acting in a way that generates the data it predicts.

The above view, of course, would still not amount to an explanation of the mechanism of reference. Furthermore, the existence of cases of reference-switching shows that we do not use such an equilibrium selection rule. Such an equilibrium selection rule has little too recommend it; it is a good idea for our semantic practice to be conservative, but such a rule would be too rigid. Still, I think that thinking about semantics in terms of determining an equilibrium selection rule has a lot to recommend it. We can, as explained earlier, distinguish between foundational theories of reference and assignment theories of reference. Assignment theories are theories that assign referents to names. On the above construal we can think of such assignment as being generated in virtue of equilibrium selection rules. If we think about assignment theories as generated in virtue of equilibrium selection rules the semanticist can use the tools of game theory to his own ends. This may be a useful way of reconceptualising semantic questions concerning cases like 'Gödel'/'Schmidt' and others.

4. Conclusion

that of a layman, which should be sufficient reason to defer to experts. (Also note that an expert's use is more likely to be stable, which may be another reason to defer to them.)

In this thesis I have claimed that the idea that Kripkean causalism explains the 'mechanism of reference' has no non-trivial content, that the traditional descriptivists confused conventional reference and supposed reference and that there is no question which is such that we can interpret traditional descriptivism and Kripkean causalism as presenting remotely plausible rival answers to it. I think, as indicated in the introduction to this thesis, that one lesson to be drawn from this is that the method of cases can be dangerously misleading. Another lesson is that we should be very careful to keep three issues distinct. These are the three issues that any theory of the conventional reference of names must be concerned with. The first issue concerns the nature of conventions. If broadly Lewisian views of convention are correct, then this is a matter of trying to determine the constraints that the nature of coordination games place on any account of the reference of names. The second issue concerns the nature of communication. If the standard view is correct, then this is a matter of trying to determine the constraints that the fact that, for communication to occur, information must go from one place to another, imposes on any possible theory of the reference of names. The third issue is a matter of trying to determine what semantic reference is, i.e. of trying to give a foundational theory of semantic reference.

Some of the constraints on any possible theory of semantic reference are set by game theory and information theory. While we should be inspired by Russell and Kripke, we should also be animated by the issues that inspired von Neumann and Shannon.

The deeper lesson to be drawn from the foregoing reasoning, however, concerns our use of language and concepts. If the claims in this thesis are accurate, then conceptual confusion has been rife in semantics[75]. The fact that this is surprising indicates that we do not have a good grasp of how conceptual confusion occurs, or possibly even of what conceptual confusion is. This, then, is the main lesson to be drawn from the foregoing reasoning: we need a theory of conceptual confusion that makes it plain what it is, how it occurs and on which the fact that semanticists have been so confused for so long is not surprising.

The idea of systematic confusion in philosophy, while currently unfashionable, is, of course, not new. Russell, Wittgenstein and the positivists all believed some version of the idea that we are prey to the 'bewitchments of language' in a way that cannot be resolved by simply being clearer about what we mean. I think that, in a deep sense, they were correct. What nor-

---

[75] Of course, we have little reason to suppose that other fields in philosophy do not have similar, though unnoticed, difficulties.

mally passes for rigour or clarity is the clear expression of what we mean. This is not the problem. Two philosophers can argue the relative merits of traditional descriptivism and Kripkean causalism, or both agree that that causal chains are the 'mechanism of reference', and understand each other perfectly. They can even use the resources of modern logic in on order to express their ideas as precisely as possible. The problem, however, is that there is no question that traditional descriptivism and Kripkean causalism can be interpreted as giving remotely plausible answers to it. Similarly, the question is not *whether* causal chains are the mechanism of reference, but *what it would be* for something to be the 'mechanism of reference'. The problem, then, with discussion on these two issues is not that people are not being rigorous or not being clear, but that, in the most literal sense, they do not know what they are talking *about*. This is a possibility that Russell, Wittgenstein and the positivists were much more attuned to.

We do not, of course, have to adopt the naïve empiricism of the positivists, or the belief that all metaphysics is meaningless, or even the idea that language is somehow to blame for such confusions. What we should do, however, is to salvage and vindicate their core insight, namely that brilliant thinkers can discuss philosophical issues over extended periods of time and still end up with views that are not even wrong[76]. Wittgenstein and the positivists viewed their task as a matter of separating sense from nonsense. We may conceptualise the project in a radically different way. But, whatever form such an enquiry may take, I hope to have shown that it deserves to once again become a central concern in philosophy.

---

[76] Note that a large number of analytic philosophers believe this anyway, they just think such people are called 'continental philosophers'. And *vice versa*.

Bibliography

Åkerman, J. 2010. Communication and indexical reference. *Philosophical Studies.* 149: 355 – 366.

Almog, J. 1984. Semantical Anthropology. *Midwest Studies in Philosophy.* IX: 479 – 489. Blackburn, Simon. 1984. *Spreading the Word.* Oxford: Clarendon Press.

Buchanan, R. 2010. A puzzle about meaning and communication. *Nous.* 44: 340 – 371.

Burge, T. 1975. On knowledge and convention. *The Philosophical Review* 84: 249 – 255.

Burge, T. 1979. Individualism and the Mental, in French, Uehling, and Wettstein (eds.) *Midwest Studies in Philosophy IV.* Minneapolis: University of Minnesota Press, pp. 73–121.

Carston, R. 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication.* Oxford: Blackwell.

Davis, W. 2005. *Nondescriptive Meaning and Reference: An Ideational Semantics.* Clarendon Press. Oxford.

Devitt, M. 1981. *Designation.* New York: Columbia University Press.

Donnellan, K. 1966. Reference and definite descriptions. *The Philosophical Review*, 75 vol. 3: 281–304.

Evans, G. 1982. (ed: Macdowell, J). *The Varieties of Reference.* Oxford University Press: New York.

Frege, G. 1948. Sense and reference. *The Philosophical Review*, 57, vol. 3: 209-230.

Frege, G. 1956. The thought: a logical enquiry. *Mind.* 65: 289 – 311.

Frege, G. 1993. Letter to Jourdain. In Moore, A. (ed). *Meaning and Reference*. Oxford University Press.

Gauker. C. 1994. *Thinking Out Loud: An Essay on the Relation between Thought and Language.* Princeton University Press.

Gigerenzer, G, Todd, M & The ABC Research Group. 1999. *Simple Heuristics that Make Us Smart*. Oxford University Press: Oxford.

Grice, H. P. 1957. Meaning. *The Philosophical Review* 66: 377 - 388.

Kaplan, D. (1989a). Demonstratives. Almog, J, et al. (eds.) 1989. *Themes from Kaplan.* Oxford University Press.

Kripke, S. 1977. Speaker's reference and semantic reference. In *Midwest Studies in Philosophy*, 2: 255-276.

Kripke, S. 1981. *Naming and Necessity.* Oxford: Blackwell Press.

Kripke, S. 2008. Frege's theory of sense and reference: some exegetical notes. *Theoria* vol. 74: 181 – 218.

Kroon, F. 1987. Causal descriptivism. *Australasian Journal of Philosophy*. 65: 1-17.

Lewis, D. 1969. *Convention: A Philosophical Study*. Cambridge: Harvard University Press.

Lewis, D. 1975. Languages and language. Keith Gunderson (ed.), *Minnesota Studies in the Philosophy of Science*, Volume VII, Minneapolis: University of Minnesota Press, 3–35.

Lewis, D. 1976. Convention: Reply to Jamieson. *Canadian Journal of Philosophy.* 6: 113 – 120.

McGinn, C. 1977. Charity, interpretation, and belief. *Journal of Philosophy*, 74: 521–535.

Miller, S. 1992. On conventions. *Australasian Journal of Philosophy*. 70: 435 – 444.

Oliver, R. 1977. *The Cambridge History of Africa, Volume 3*. Cambridge University Press.

Putnam, H. 1975. The Meaning of 'Meaning'. *Mind, Language and Reality: Philosophical Papers: Vol. 2 - Mind, language and Reality*. Cambridge: Cambridge University Press.

Raatikainen, P. 2006. Against causal descriptivism. *Mind and Society*. 5: 78 – 84.

Reimer, Marga. Reference. *The Stanford Encyclopedia of Philosophy (Spring 2010 Edition)*. Edward N. Zalta (ed.) http://plato.stanford.edu/archives/spr2010/entries/reference/.

Russell, B. 1905. On Denoting. *Mind*, New Series. 14: 479–493

Russell, B. 1910. Knowledge by acquaintance and knowledge by description. *Proceedings of the Aristotelian Society.*11: 108 – 128.

Russell, B. 1957. Mr. Strawson on referring. *Mind*. 66: 385 – 389.

Russell, B. 2009. *The Philosophy of Logical Atomism*. Routledge: London.

Sainsbury, M. 2002. Russell on names and communication. In Sainsbury, M. 2002. *Departing from Frege.* Routledge: London.

Salmon, N. 1986. *Frege's Puzzle*. Cambridge: MIT Press.

Schelling, Thomas. 1960. *The Strategy of Conflict*. Cambridge: Harvard University Press.

Searle, J. 1983. *Intentionality: an Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

Simon, H. 1997. *Models of Bounded Rationality. Volume 3: Empirically Grounded Economic Reason*. Cambridge: MIT Press.

Shannon, C. 1948. A mathematical theory of communication. *Bell System Technical Journal*. 27: 379–423

Soames, S. 2002. *Beyond Rigidity.* New York: Oxford University Press.

Smit, J. 2012. Why bare demonstratives need not semantically refer. *Canadian Journal of Philosophy.* 42: 43 – 66.

Speaks, J. Theories of meaning. *The Stanford Encyclopedia of Philosophy (Summer 2011 Edition).* Edward N. Zalta (ed.) http://plato.stanford.edu/archives/sum2011/entries/meaning/

Strawson, P. 1950. On referring. *Mind.* 59: 320 – 344.