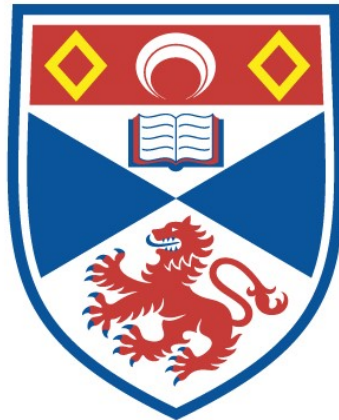


VEHICLE RELATIONISM: ESSAYS ON SAMETHINKING AND  
SAMESAYING

Sara Kasin Vikesdal

A Thesis Submitted for the Degree of PhD  
at the  
University of St Andrews



2019

Full metadata for this item is available in  
St Andrews Research Repository  
at:  
<http://research-repository.st-andrews.ac.uk/>

Please use this identifier to cite or link to this item:  
<http://hdl.handle.net/10023/18564>

This item is protected by original copyright

# Vehicle Relationism: Essays on Samethinking and Samesaying

Sara Kasin Vikesdal



University of  
St Andrews

This thesis is submitted in partial fulfilment for the degree of

Doctor of Philosophy (PhD)

at the University of St Andrews

January 2019



### **1. Candidate's declaration**

I, Sara Kasin Vikesdal, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 65 000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews and the University of Oslo in November 2015 as part of a joint programme.

I confirm that no funding was received for this work.

Date ..... signature of candidate .....

### **2. Supervisor's declaration**

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date ..... signature of supervisor .....

### **3. Permission for publication**

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

I, Sara Kasin Vikesdal, confirm that my thesis does not contain any third-party material that requires copyright clearance.

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

PRINTED COPY

a) No embargo on print copy

ELECTRONIC COPY

a) No embargo on electronic copy

Date ..... signature of candidate .....

Date ..... signature of supervisor .....

#### **4. Underpinning Research Data or Digital Outputs**

##### **Candidate's declaration**

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

Date ..... signature of candidate .....





# *Acknowledgements*

Many people have helped me during the process of writing this thesis, and for that I am very grateful.

First and foremost I wish to give a special thanks to my supervisors, Derek Ball and Herman Cappelen, for their encouragement and invaluable feedback on the material in this thesis. I would also like to thank my second supervisor, Josh Dever, for helpful discussions along the way.

Thanks are also due to Solveig Aasen, Torfinn Huvenes, Ole Martin Moen, Andrea Onofri, Simon Prosser, Francois Recanati, Mark Sainsbury, Margot Strohminger, Sebastian Watzl and Juhani Yli-Vakkuri for helpful feedback and discussion. Thanks to my fellow PhD students in Oslo and St Andrews. In Oslo this includes Mirela Fuš, Sigurd Jorem, Max Kippersund, Joonas Räsänen, Monica Roland, Jeroen Rijnders, Eduardo Saldaña, Maria Seim, Aksel Braanen Sterri, Kristoffer Sundberg, Leo Townsend, Joost Vecht, Hilde Vinje, Yelena Yermakova, Lars Gjoevikli and Jola Feix. I'm also grateful to the students at Arché for helpful discussions: Marvin Backes, Miguel Egler, Claire Field, Dan Healey, Ethan Landes, Quentin Pharr, Lewis Ross, Alessandro Rossi, Ravi Thakral, Caroline Touborg and Alex Yates. A special thanks to my office mates, Lisa Bastian and Anh-Quan Nguyen. Thanks to the members of ConceptLab and CSMN.

Lastly, I would like to thank my family, in particular my parents, Helle and Inge, for their love and encouragement, and Kim and Mik for an unbelievable amount of support and help in writing this thesis. I could not have done this without you.





## **Table of Contents**

Introduction	3
De Jure Co-Reference: In Defence of Pointers	53
Vehicle Relationism: In Defence of Pointers	81
Originalism and Coordination in Thought: In Defence of Vehicle Relationism	131
Staying on Topic: The Continuance-of-Topic Relation is Non-Transitive	163
Bibliography	191



# Introduction

---

This thesis is about the nature of *samethinking* and *samesaying*. These notions are broad and capture various distinct but related phenomena. I will focus on two particular understandings of ‘samethinking’, and on one particular understanding of ‘samesaying’. Let me address samethinking first. On the first understanding of ‘samethinking’, samethinking occurs whenever two thoughts concern the same referent.<sup>1</sup> We may distinguish between two different ways in which this can occur. First, there are cases in which the sameness of reference is manifest to the subject. Take for instance the two beliefs BOB DYLAN IS A MUSICIAN and BOB DYLAN WON A NOBEL PRIZE.<sup>2</sup> In such cases, the sameness of reference is transparent to the thinker in such a way that she may combine the two beliefs in an inference and conclude directly from these two beliefs alone that a musician won a Nobel Prize.

Second, there are cases in which two thoughts concern the same referent, but where the sameness of reference is not manifest to the subject. Take for instance the two thoughts BOB DYLAN IS A MUSICIAN and ROBERT ZIMMERMAN WON A NOBEL PRIZE. ‘Robert Zimmerman’ is Bob Dylan’s birth name, so the two thoughts concern the same individual. However, unless the thinker has a further belief to the effect that Bob Dylan is Robert Zimmerman, she may not rationally infer from these beliefs that a musician won a Nobel Prize. We see, then, that two pairs of thoughts that are referentially equivalent may nonetheless play different roles in cognition. In this thesis, I offer a novel account of how to understand the difference between cases of samethinking such as those above.

---

<sup>1</sup> By ‘thoughts’, I mean psychologically instantiated mental representations.

<sup>2</sup> Throughout the thesis, I use capital letters to indicate concepts and thoughts.

The second understanding of ‘samethinking’ that I will discuss in this thesis is a broader phenomenon. Two thoughts, typically entertained by distinct individuals or the same individual at different times, can be said to concern the same subject matter despite differing in their overall semantic properties. Likewise, it seems that two utterances may concern the same topic despite differing in their overall semantic properties. Consider for instance someone uttering the sentence “Whales are fish” in the 18<sup>th</sup> century, where such an utterance would generally be regarded as true. If someone today were to utter the same sentence, however, we would regard it as false. We have reason to think that the meaning of the term ‘fish’ has changed between then and now. Even if this is the case, it seems as though the 18<sup>th</sup>-century person and the current day individual are, in an interesting way, talking about the same topic when uttering the sentence.<sup>3</sup> This is the notion of ‘samesaying’ I will address in this thesis. I shed light on what it is for two thoughts or two utterances to be the same in this way.

The thesis consists of four chapters that are to be seen as individual papers. I have, however, included cross-references between the chapters where this is appropriate.<sup>4</sup> The first three chapters focus on the first interpretation of ‘samethinking’: they are concerned with the distinction between cases in which co-reference is manifest to thinkers and cases where it is not. The general question I engage with is how we should understand the nature of thought in order to account for the sort of rational reasoning that hinges on such samethinking. I develop a new framework for understanding the nature of thoughts. This framework accounts for the difference between cases of samethinking in which sameness of referent is manifest and cases where it is not in terms of primitive representational relations. I use this framework to account for some of the long-standing puzzles within the philosophy of mind, such as Frege’s Puzzle, and argue that the suggested framework is superior to central alternative views on the nature of thought.

---

<sup>3</sup> Whether or not cases such as ‘fish’ actually involve a change in meaning is controversial. I say more about this in chapter 4. For now, I use the case as an example in order to provide an intuitive grip on the relevant notion.

<sup>4</sup> Since the chapters are to be seen as individual papers, there is some overlap of content in certain places.

The final chapter of this thesis concerns the broader understanding of ‘samethinking’ as well as the corresponding notion of ‘samesaying’. I discuss how there can be stability of topic in cases where there is a change in the semantic properties of the relevant representational devices. I argue that the continuance-of-topic relation is non-transitive and that this puts restrictions on an account of stability of topics.

In this introductory chapter, I will give an overview of the thesis and present some background for the general discussion. In section 1, I present the notion of *coordination*, which is central to the discussion of samethinking in the first three chapters. In section 2, I briefly present the historical background for this discussion. In section 3, I introduce the central competing views with which I engage in this thesis. In section 4, I present the key claims of my positive proposal, which I call *Vehicle Relationism*. In section 5, I lay out the background for the discussion of stability of topics, which is the central theme of chapter 4. In section 6, I give an overview and summary of each of the thesis chapters. Finally, in section 7, I point to some of the main findings of this thesis as well as suggestions as to future work based on these findings.

### 1. Coordination in Thought

Our dispositions to act depend on our minds combining information in specific ways. In particular, the mind treating certain pieces of information as concerning the same referent is essential for such behavioural dispositions. If, for instance, I have a desire concerning a particular individual and also certain beliefs concerning this specific individual I may combine such propositional attitudes and act rationally on them only if my cognitive system treats these propositional attitudes as concerning the same individual. Further, rational reasoning depends on the possibility of the beliefs figuring as premises in an inference being treated as concerning the same referent. An important question is this: What is the nature of thought such that our minds may combine information in this way?

Whenever two mental representation tokens are so related that the subject may combine the relevant information in inferences, or so as to act on them, *directly*, without making an explicit identity judgement, we say that the representations are *positively coordinated*.<sup>5</sup> In contrast, if the thinker is not rationally disposed to exploit the relevant information in this way the representations in question are *negatively coordinated*. In this section, I will elaborate on the phenomenon of coordination and show why understanding the nature of this relation is an important philosophical undertaking.

### *1.1. Coordination and Behavioural Dispositions*

Suppose you have a desire to learn more about Bob Dylan. Suppose also that you have the belief that there are several books about Bob Dylan in your local library. You also form the belief that the best way for you to learn more about Bob Dylan is to go to this library and read about him. You then have a reason to go to the library and read certain books there.<sup>6</sup>

In order for you to be able to act on such reasons your cognitive system must treat the relevant beliefs as concerning the same individual as the person you wish to learn more about. In other words, the fact that these propositional attitudes are treated as concerning the same individual is essential to your behavioural dispositions.

To see this, contrast this scenario with one in which you have a desire to learn more about Robert Zimmerman. Since 'Robert Zimmerman' is Bob Dylan's birth name, this desire concerns the same individual as in the previous scenario. However, let's imagine that you are not aware of this fact. You have

---

<sup>5</sup> The term is originally due to Fine (2007). Fine sometimes uses the term in a theory-neutral way and sometimes as a term integrated in his specific framework. I follow Gray (2017, 2018) in using 'coordination' in a theory-neutral way. 'Coordination', in this sense, "just picks out a rationally relevant relation between representations" (Gray, 2018, 2 n.3). As we'll see later in this section, the rationally relevant relation in question is the one that figures into the explanation of the cognitive system treating pieces of information as concerning the same referent in such a way as to warrant trading on identity and to yield certain behavioural dispositions.

<sup>6</sup> I assume here a simple belief/desire account of agency (cf. Davidson 1963). Nothing hangs on this.

independently come to know about Robert Zimmerman through a mutual friend who has told you many stories about what it was like growing up next door to Robert Zimmerman. Let's keep the beliefs in the initial scenario fixed so that you still believe that the best way for you to learn more about Bob Dylan is to read books about him at the library. In this case, since you are not aware of the fact that your desire to learn more about Robert Zimmerman concerns the same individual as your beliefs regarding how best to learn more about Dylan, you do not have the behavioural disposition to go to the library based on this set of propositional attitudes alone. In this case, the cognitive system does not combine the relevant propositional attitudes in the way it did in the first scenario. Although the different attitudes in fact concern the same individual, the cognitive system treats the information as concerning distinct referents. In this case, you need to form a further belief concerning the identity of Zimmerman and Dylan in order to have a (rational) behavioural disposition to go to the library based on this set of attitudes.

What aspects of thoughts explain the difference between the sets of propositional attitudes in the two cases? It cannot be the referential content, since the referent is the same in both cases. In the first scenario, the sameness of reference of your beliefs and desire is manifest to you, whereas in the second case the sameness in reference is not manifest to you. This marks the main difference between the two cases. The way in which you think about the individual seems to make a difference to whether or not the sameness in reference is manifest to you. Whether or not the sameness of reference is manifest, in turn, explains the difference in your behavioural dispositions in the two cases. The first scenario illustrates a case in which your beliefs and desire are positively coordinated. The second scenario illustrates a case in which your beliefs and desire are negatively coordinated. In order to rationally act on your belief/desire pair in the first scenario, you do not need to make any explicit identity judgement to the effect that the attitudes concern the same individual. You simply combine these attitudes directly as a result of the sameness of reference being manifest to you. This phenomenon is closely related to that of *trading on identity*, to which I now turn.



1.2. *Trading on Identity*

The ability to combine pieces of information taken to concern the same referent is essential to the explanation of rational reasoning. If you believe that Bob Dylan is a musician and also that Bob Dylan won a Nobel Prize you may combine this information and rationally infer that a musician won a Nobel Prize. This is because your two beliefs are positively coordinated. If, in contrast, you believe that Robert Zimmerman was a musician and that Bob Dylan won a Nobel Prize, your beliefs would be negatively coordinated and as a result, you would not be rational in inferring from these two beliefs alone that a musician won a Nobel Prize. In order to be rational in making this latter inference, you would also need to have a further belief to the effect that Zimmerman is Dylan. What is the nature of the state you're in when you are warranted in making the inference without making an explicit identity judgement?

As I said, in cases where your beliefs are negatively coordinated, the inference in question would not be warranted, since you do not believe that Zimmerman is Dylan. At first glance, then, it might seem intuitively plausible that the reason why you are rationally warranted in making the inference in the first scenario – when your beliefs are positively coordinated – is that in this situation you *do* make such an identity judgement. This judgement, then, would result in a new belief to the effect that Bob Dylan is Bob Dylan. On this picture, this is what the two inferences would look like:

BOB DYLAN IS A MUSICIAN	ROBERT ZIMMERMAN IS A MUSICIAN
BOB DYLAN WON A NOBEL PRIZE	BOB DYLAN WON A NOBEL PRIZE
[BOB DYLAN IS BOB DYLAN]	[ROBERT ZIMMERMAN IS BOB DYLAN]
A MUSICIAN WON A NOBEL PRIZE	A MUSICIAN WON A NOBEL PRIZE

The suggestion under consideration, then, is that the difference between the two cases is due to the fact that, in the first case you do have the bracketed belief whereas in the second case you do not have the bracketed belief. If this were

true, the difference between positive and negative coordination could be explained in terms of such implicit identity judgements obtaining or failing to obtain.

There are, however, serious problems with understanding the difference between the two cases in this way. Notice that the bracketed beliefs contain concept tokens that must be recognized as concerning the same individual as the concept tokens in the prior premises. That is, the singular concepts in the bracketed beliefs (i.e. [BOB DYLAN and BOB DYLAN] in the first inference and [BOB DYLAN and ROBERT ZIMMERMAN] in the second inference) must be positively coordinated with the corresponding singular concepts in the preceding premises. How do we account for the recognition of sameness in reference in this case? If, as we have been assuming, we need identity judgements in order to recognize the sameness of reference, we would have to introduce further identity premises in order to account for the singular concepts in the bracketed beliefs being treated as concerning the same individual as the singular concepts in the preceding premises. But we would then run into further problems when trying to explain how one can recognize that these further identity judgements concern the same individual as the other beliefs. And the same problem would arise again: an infinite regress ensues.

We see, then, that the recognition of sameness in the first inference, and hence someone being rational in making such an inference, cannot be due to further identity judgements to the effect that the singular concepts refer to the same individual. Instead, the correct way to understand the inference is thus:

BOB DYLAN IS A MUSICIAN
BOB DYLAN WON A NOBEL PRIZE
<hr style="width: 30%; margin: 0 auto;"/>
A MUSICIAN WON A NOBEL PRIZE

There is no need to include the identity judgement in order to account for why someone may be rational in drawing this inference. The sameness of reference of

the singular concepts in the premises is directly manifest to the thinker. This is what sets the two inferences apart: In the case of someone having thoughts about Robert Zimmerman and Bob Dylan without realizing that they are the same person, they must make the further discovery that Dylan is Zimmerman in order to rationally draw the inference, whereas in the other case the sameness of reference is simply manifest in such a way that no further identity judgement is needed for the inference to be rational. This marks the key difference between those beliefs that are positively coordinated and those that are negatively coordinated.

When sameness of reference is manifest to thinkers and the thinker exploits this manifestness in drawing inferences in this way, we may say that the thinker is *trading on identity of co-reference*. The notion of 'trading on identity' is originally due to Campbell (e.g. 1987, 1994, 2002, 2012). He considers a deductive argument of the same form as the inference concerning Bob Dylan above, where the singular terms in each premise is manifestly co-referential to the thinker without the need for an implicit identity judgement:

This argument is valid as it stands. There is no need for an extra premise asserting the identity of [Dylan] (as referred to in the first premise) with [Dylan] (as referred to in the second premise). If you did suppose that such a premise is needed, and provided it, you would have only begun. For you would now need further premises asserting the identity of all the various [Dylans] referred to in the course of the inference; and no finite, or for that matter, infinite provision of further premises would be enough. We have to acknowledge that the argument is valid as it stands. Rather than depending on an implicit identity premise connecting the terms of the first two premises, it simply 'trades on' the identity of reference of those terms. (Campbell 2012, 97)

I take this to be a convincing reason to deny that trading on identity depends on implicit identity premises. That is, it cannot be the case that positive

coordination obtains in virtue of such identity judgements. How then are we to understand what warrants trading on identity?

We need a better understanding of the nature of coordination relations. This is the central task with which I will be concerned in the first three chapters of the thesis. I will consider some of the most prominent contemporary accounts of coordination relations and offer criticisms of such views. I will propose a new understanding of coordination relations that accounts for the phenomenon of trading on identity. This framework draws on some key insights from previous suggestions found in the literature, but combines these insights in a new way. The result is a novel account of coordination.

I will say more about this later in this chapter, but first I will present some historical background for the debate. Although most of the views I engage with in the thesis are contemporary, the interest in the phenomenon of trading on identity and related phenomena goes back to at least Frege (e.g. 1892) and Russell (e.g. 1910, 1912). The various accounts of coordination proposed in the current debate are in many ways informed by the various views on the nature of thoughts stemming from Frege and Russell. There is in particular one observation made by Frege that marks the outset of the debate, namely his (1892) insight that sameness of reference is not always manifest to thinkers.

## 2. Historical Background

Contemporary accounts of the nature of thoughts fall into two broad categories: On the one hand there are those who have a broadly Fregean view on the nature of thoughts, and on the other, there are those who have a broadly Russellian view on the nature of thoughts. Both camps maintain that thinking involves standing in a relation to propositions, but they differ in what they take the nature of such propositions to be.<sup>7</sup> In what follows, I will point to the key distinguishing

---

<sup>7</sup> For the sake of simplicity, I set aside views that take propositions to be sets of possible worlds and views that take propositions to have no structure. My positive account does not say anything about the nature of propositions. I do, however, accept a structured view on propositions. I return to this later on.

aspects of such theories. I will also point to central puzzles found in the literature that form the basis for the current debate.<sup>8</sup>

### 2.1. Fregeanism

The first broad group of views accepts Frege's (1892) distinction between *sense* and *reference*.<sup>9</sup> According to such views, thoughts about objects in the external world are not directly about these objects, but instead they are mediated by senses associated with such objects. This view is closely connected to the thesis that thinkers are only related to objects in the external world via the properties of such objects. When we think about objects, we do not think about them directly, but indirectly, via their properties. Senses can thus be understood as descriptive satisfaction conditions that determine the referent of concepts and thoughts.<sup>10</sup> So, for instance, the reference of a singular concept such as BOB DYLAN is Dylan himself, while the sense of this concept is a way of thinking about Bob Dylan, a *mode of presentation* of Dylan.

According to Frege, the sense of a thought is a proposition.<sup>11</sup> Each constituent concept of a thought has a sense and together they combine into a complex sense of the thought as a whole. The reference of a thought is its truth-value. A thought's truth-value depends on the reference of the constituent concepts and such reference is determined by the senses of the concepts. On this view, there

---

<sup>8</sup> Frege and Russell exegesis is beyond the scope of this thesis. In what follows I present the relevant views in accordance with standard contemporary presentations of such views (e.g. Recanati 2012, Chapter 1; Sainsbury & Tye 2012, Chapter 2). I will also present the views in such a way that their relevance to the contemporary debate about coordination becomes clear, rather than going into details about Frege and Russell's views that are not directly relevant to the discussion in this thesis.

<sup>9</sup> More specifically, this broad category of views takes senses to be constituents of propositions (e.g. Evans 1982, McDowell 1994, Stanley 2011; c.f. Merricks 2015, 129).

<sup>10</sup> Some philosophers (e.g. Burge 1979, Evans 1982, McDowell 1984) have argued that Fregean senses need not be construed exclusively as descriptive senses. Instead, we may allow for non-descriptive or *de re* senses. I put this possible complication aside in this presentation since nothing of what I say depends on it. I will, for simplicity, follow Kripke (1980) and assume that Frege was a descriptivist.

<sup>11</sup> Frege uses the word 'thought' to denote what is commonly called 'propositions'. When presenting Frege's view I stick to standard terminology rather than Frege's own terminology. Hence, I use 'thought' in the ordinary sense rather than in Frege's technical sense. I will use the term 'proposition' when talking about what Frege called 'thoughts'.

are two components to the content of concepts and thoughts; their senses and their reference. Thinkers are only directly related to the former.<sup>12</sup>

Frege's central argument in favour of positing senses is tightly bound up with the phenomenon of samethinking, discussed in the previous section. Frege argues for the distinction between sense and reference in the following way. Let's assume that there is nothing more to the content of atomic concepts than their referents. If this is the case, the content of BOB DYLAN is just Dylan himself.<sup>13</sup> Similarly, the content of ROBERT ZIMMERMAN is also Dylan himself. On this view, then, DYLAN and ZIMMERMAN have the same content. Further, given plausible assumptions about compositionality, the proposition that Dylan won a Nobel Prize and the proposition that Zimmerman won a Nobel Prize is the same. But as we saw in the previous section, this seems counterintuitive. One would have to explain how someone who fails to realize that DYLAN and ZIMMERMAN are co-referential could rationally believe and reject the very same proposition. If thinking is a direct relation between thinkers and propositions, how can it be that someone may rationally accept and deny the very same proposition?

Further, anyone who has the concept DYLAN knows a priori that Dylan is Dylan. However, if all there is to the contents of atomic concepts are their referents, the proposition that Dylan is Dylan must be the same as the proposition that Dylan is Zimmerman. But it seems intuitively plausible that not everyone who knows that Dylan is Dylan knows that Dylan is Zimmerman. In fact, the latter carries potentially valuable information that allows someone who comes to know it to gain new knowledge about the world. The thought that Dylan is Dylan and the thought that Dylan is Zimmerman thus seem to play different roles in cognition: The former is trivial while the latter is potentially informative. It seems that if

---

<sup>12</sup> C.f. Carruthers' (1989) presentation of Frege's view.

<sup>13</sup> Frege, of course, does not use Dylan as an example. Instead Frege talks about the Ancient Babylonians. In Ancient Babylonia people called the brightest star visible in the morning 'Phosphorus'. They gave the brightest star visible in the evening the name 'Hesperus'. Unbeknownst to them 'Hesperus' and 'Phosphorus' refer to the same heavenly body, namely the planet we now call 'Venus'. For the sake of consistency, I will present Frege's puzzle in terms of BOB DYLAN and ROBERT ZIMMERMAN. However, the puzzle as portrayed here is analogous to Frege's initial puzzle.

the two thoughts express the same proposition we have no way of explaining how they may nonetheless differ in cognitive significance.

This case shows that if thinking is a direct relation between individuals and propositions, the contribution of a singular concept to a proposition cannot simply be its referent. This is because two thoughts that are referentially isomorphic may potentially play different roles in cognition. One way to cash out the implication of Frege's observation is in terms of coordination (c.f. Gray 2017). Coordination, remember, is tied in with the phenomenon of manifest co-reference. Since rational individuals may fail to know that DYLAN and ZIMMERMAN are co-referential such concept tokens are negatively coordinated. Since the two concepts are in fact co-referential, coordination cannot merely be a matter of having the same reference. We need something more in order to explain the possibility of someone recognizing or failing to recognize the sameness of reference.

According to Frege, senses are what accounts for the cognitive role of concepts and thoughts. Even though two concept tokens have the same reference they may have different senses. This means that, on Frege's view, two thoughts or concepts that play different roles in cognition have different senses. As a result, two thoughts that differ in their cognitive role differ in their propositional content. If two thoughts differ in truth-value, or if it is rationally permissible for a subject to take conflicting attitudes toward their contents, the two thoughts have different senses. This means that they have distinct propositional contents.

Campbell (1987) explicitly uses senses to account for coordination:

Any account of logical form which deals with inferences depending upon two occurrences of singular terms having the same referent will need to say when one can, as I shall put it, simply *trade upon* the fact that they have the same referent. [...] it is in dealing with [this] that we see why we need the notion of the *sense* of a singular term. (Campbell 1987, 275)

For Fregeans, positive coordination is explained in terms of sameness of sense whereas negative coordination is explained in terms of a difference in sense. That is, whether or not trading on identity is warranted depends on whether or not there is sameness of sense.

Understood this way, Frege's puzzle is ultimately a puzzle about coordination. I thus accept what Gray (2017) calls *Cognitive Significance as Coordination*, which is the claim that "differences in cognitive significance between representations with the same referential content are explained by coordination" (2017, 3). The key aspect of Fregeanism about coordination is this: Whether or not two concept tokens are positively coordinated is to be explained in terms of sameness or difference in senses. I now turn to Russellianism.

## 2.2. Russellianism

The other broad category of accounts of the nature of thoughts and propositions that I will consider is the Russellian view. Russellians follow Russell in rejecting Frege's distinction between sense and reference.

According to Russell (e.g. 1903) propositions are structured complexes of objects and relations. He says that

Whatever may be an object of thought, or may occur in any true or false proposition [...] I call a term. [...] A man, a moment, a number, a class, a relation, a chimaera, or anything else that can be mentioned, is sure to be a term; and to deny that such and such a thing is a term must always be false. (Russell 1903, 43)

Propositions that have singular individuals or objects as constituents are called *singular propositions*.

Russell held that a subject must stand in a direct relation (what he calls an *acquaintance* relation) to all the constituents of a thought in order for the subject to be able to entertain it (1910, 1912). In his early days, Russell thought that it is



possible for individuals to have thoughts concerning mind-external entities directly, as illustrated by the quote.<sup>14</sup> However, in his later works he holds that the only sorts of things that thinkers can be acquainted with are their own occurrent sense data, universals and, perhaps, also the self:

We have acquaintance in sensation with the data of the outer senses, and in introspection with the data of what may be called the inner sense—thoughts, feelings, desires, etc.; we have acquaintance in memory with things which have been data either of the outer senses or of the inner sense. Further, it is probable, though not certain, that we have acquaintance with Self, as that which is aware of things or has desires towards things. (Russell 1912, 51)

As a result, the only singular propositions human beings are capable of having propositional attitudes towards are those that have such entities as constituents.

Russell's reason for thinking that we cannot stand in direct relations to mind-external entities is similar to Frege's reason for introducing the notion of a sense. Russell thought that one could only be acquainted with that for which misidentification is not possible. If someone could be presented with the same object twice and rationally fail to realize that it is the same object, then they do not stand in an acquaintance relation to that object. If it is possible for two thoughts to concern the same object but where the co-reference is not manifest to the thinker, the thinker is not acquainted with the relevant object. In the case of BOB DYLAN and ROBERT ZIMMERMAN, we see that a thinker may fail to realize that two thoughts, that contain each concept respectively, concern the same individual. Thus, the thinker is not acquainted with Bob Dylan. Russell agrees with Frege that the content of propositions is transparent to thinkers in that thinkers cannot fail to recognize sameness. Hence, there must be a difference in the content of BOB DYLAN and ROBERT ZIMMERMAN that accounts for the possibility of misidentification.

---

<sup>14</sup> To give another illustration: In a letter to Frege, he famously said that "I believe that in spite of all its snowfields Mont Blanc itself is a component part of what is actually asserted in the proposition 'Mont Blanc is more than 4,000 metres high'" (Letter to Frege, 12 December 1904, in Frege 1980, 169).

The resulting view is one according to which our thoughts about mind-external objects are never direct, but always mediated by our sense data, to which we do stand in direct relations. If you have a thought about Bob Dylan, it is not directly Dylan himself you have a thought about. Rather, your thought about Dylan is of the form *the thing that caused THIS [referring to your occurrent sense datum] is so-and-so* (c.f. Russell 1910). Our knowledge about the external world, then, is grounded in acquaintance relations between ourselves and sense data. Whenever someone has a thought about determinate singular objects in the external world they deploy a singular thought. But according to Russell, it is not the mind-external object that is a constituent of the singular proposition, but rather it is a sense datum.

This contrasts with the Fregean doctrine that all thoughts are mediated by (Fregean) senses and that one can never think about anything, not even sense data, in an unmediated way. For Frege, there are no singular propositions. In short, we may say that the main difference between Fregean and Russellian propositions is that the former are composed wholly of senses, while the latter have particulars (such as particular sense data) as constituents.

In current debates, many who are sympathetic to Russell's general framework, which takes propositions to be structured complexes of objects and relations and denies the need for a further level of semantic content such as senses, resist Russell's restrictivism about acquaintance relations and singular thoughts. They deny the sort of descriptivism advocated by Russell (1912). Such philosophers maintain that one may have singular thoughts about mind-external objects as well. Hence, according to such views, your thought that Bob Dylan won a Nobel Prize contains Dylan himself as a constituent, rather than sense data caused by that individual. I will call this view of propositions *Millian/Russellianism*. The Millian aspect of the view comes from Mill's (1843) claim that the referent of a proper name is just its referent. Transposed to the realm of thought, the view holds that the semantic content of singular concepts is their referent. Singular concepts are those concepts that refer to individual objects in an unmediated way (i.e. not by way of descriptions etc.). The Millian claim, then, is that singular concepts may refer directly to mind-external particulars.

A common feature of Frege's framework and Russell's framework is that they take thinking to be a direct relation between subjects and propositions. We may call this the *Dyadic View of Thoughts*. Many of those who take propositions to be Millian/Russellian nowadays typically deny this; instead they take thinking to be a triadic relation between subjects, mental representations and propositions. I will say more about this approach in section 3. Before that, let me briefly set out how the discussion in this thesis will relate to the general debate between Fregeanism and Russellianism.

In this thesis I set aside the general question of which of the broad views of the nature of thoughts is correct. However, the general aim of the proposed account of coordination can be seen as a contribution to the debate in that, if successful, it provides an argument in favour of Millian/Russellianism. Although, as I point out throughout the thesis, the proposed framework of coordination is, in theory, compatible with any account of the nature of propositions, I will in discussion assume a Millian/Russellian account. My overarching aim is to see whether it is possible to do justice to the Fregean data that co-referential concepts may differ in their cognitive significance, within a broadly Millian/Russellian framework. Hence, I will not discuss Fregean views at length. The views I engage with and that I take to be my main opponents are views that share many fundamental assumptions with the proposed framework. I will set out these views in section 3 below. All of these views are developed at least partly with the aim of accounting for the Fregean data within a broadly Millian/Russellian framework.

As noted, if my account of coordination is successful this provides an argument in favour of the Millian/Russellian picture. The Millian/Russellian framework is more metaphysically parsimonious than the Fregean framework, in that it only requires a one-levelled semantics. Also, various serious objections to the Fregean view have been raised in the literature. Among these, Kripke's (1979) puzzle of Paderewski is most relevant to the discussion about coordination in thought. I present the puzzle several places in the thesis, but the main observation that poses problems for the Fregean view is this: There are cases where sameness or difference in sense is not manifest to thinkers. If this is correct, senses cannot account for coordination since it would be possible for the thinker to misidentify

senses. I aim to explain both the Fregean data and the observation made by Kripke within a Millian/Russellian framework. In this way, the arguments in this thesis lend support to the Millian/Russellian picture of the nature of the content of thoughts.

Since I presuppose a Millian/Russellian framework I will not say much about the nature of the content of mental representation throughout the thesis. In fact, strictly speaking, I don't think anything about the semantic content of thoughts follows directly from my view of coordination. However, as I said, I do think that if my positive framework can account for the Fregean data without having to abandon Millian/Russellianism, this gives us reason to prefer Millian/Russellianism since overall the view is the most parsimonious.

Note that Frege's observation that co-referential terms may differ in their cognitive significance gives rise to several related but nonetheless distinct puzzles. One question that arises from the Fregean data concerns the semantics of belief ascriptions. Consider the following two ascriptions:

- 1) Lois Lane believes that Clark Kent is Clark Kent
- 2) Lois Lane believes that Clark Kent is Superman

Many hold that such belief ascriptions may differ in truth-value despite both attributing a belief to Lois Lane that involves an identity statement between the same individual.<sup>15</sup>

Further, it seems that two natural language sentences such as

- 3) Clark Kent is Clark Kent
- 4) Clark Kent is Superman

differ in informativeness: The former seems trivial, whereas the latter is potentially informative. How can this be, if the two sentences have the same semantic content?

---

<sup>15</sup> E.g. Crimmins & Perry (1989), Richard (1990).

I will not address questions pertaining to the semantic content of sentences in natural language. Rather, my focus when it comes to coordination is on thoughts and the explanation of rational reasoning and behavioural dispositions. Hence, I will not explicitly engage with the linguistic versions of the puzzle in this thesis. Some of what I say may be relevant to understanding coordination in language, but the nature of thoughts and natural language differ to such an extent that giving an account of coordination in both domains would be too ambitious for a single thesis.<sup>16</sup>

As a result, I do not engage with Millian/Russellian frameworks that focus primarily on accounting for coordination in language. The theories I have in mind are those that go under the label *Descriptive Millianism*. This is the view that, “although sentences that contain names express singular propositions, when they use those sentences speakers communicate descriptive propositions” (Caplan 2007, 181).<sup>17</sup> Such theories account for the Fregean data in language in terms of pragmatic considerations. Such pragmatic considerations are, however, not easily transferable to thoughts. For similar reasons, I do not in this thesis consider Millian/Russellian views that focus on explaining the Fregean puzzle pertaining to belief ascriptions in terms of a semantic difference. According to such views, propositional attitude reports involve unarticulated constituents that concern the way in which the thinker believes what she believes.<sup>18</sup> Although I do not engage with the debate about coordination in language, I do address a related issue pertaining to language in the final chapter (Chapter 4). There I discuss how utterances may concern the same subject matter despite a change in meaning. I say more about the background for this particular discussion in

---

<sup>16</sup> I do, however, say more about the relation between coordination in thought and language in section 7.4. below.

<sup>17</sup> Soames (2002) provides the fullest defence of Descriptive Millianism. See Barber (2000) for a similar view. For a criticism of Descriptive Millianism, see Caplan (2007). For a response to Caplan’s argument against Descriptive Millianism, see Speaks (2010).

<sup>18</sup> See, for instance, Crimmins & Perry (1989) and Crimmins (1992) for a defence of this version of Millian/Russellianism. Such views presuppose a representational theory of mind in that their explanation of the semantics of belief ascriptions essentially involves reference to such representations. Although I do not engage with this specific account of coordination in the case of the semantics of belief ascriptions, I do consider Perry’s (1980) general view of mental representations as mental files (see section 3.2. below).

section 4 below, but first I will set out the theories I discuss in the first three chapters of the thesis.

### 3. Competing Views

The views I engage with in this thesis are current prominent accounts of coordination in thought that share certain fundamental assumptions with my positive account. These views include Fodor's (e.g. 1975, 2008) Language of Thought Hypothesis (LOTH), the Mental File Framework (c.f. Recanati 2012, 2016), Originalism (Sainsbury & Tye 2011, 2012) and Semantic Relationism (Fine 2007). The first three views take thinking to be a triadic relation between subjects, mental representations, and propositional content. On such views, thinkers are not directly related to the propositional content of their thoughts. Rather, they stand in direct relations to mental representations that have such propositional contents. As we'll see, this allows us to account for the problems raised by Frege and Russell in terms of something other than sameness or difference in the propositional content of thoughts. The last view, Semantic Relationism, does not appeal to a distinction between mental representations and content. This framework, however, bears certain other similarities to the view developed in this thesis; both views take coordination in thought to be accounted for in terms of relational features of such entities. In this section, I will give a brief overview of each framework in turn. I do not, of course, intend the following presentations to cover the full extent of these complex views. Rather, I will focus on those aspects of the theories that will be important for the discussion of coordination in this thesis.

The views I will consider can be divided into two broad categories based on how they account for coordination relations. On the one hand, there are what I call *intrinsicist* views, and on the other, there are *relational* views of coordination. Before going into the details of specific frameworks, let me say a few words about these general categories.

### 3.1. *Intrinsicism and Relationism*

We may distinguish between two classes of views about how to account for coordination in thought. On the one hand, there are those who hold that coordination is to be accounted for in terms of *intrinsic representational properties* of concepts and thoughts. I borrow this notion from Gray (2017, 2018). It is to be understood as a technical notion. The properties in question need be neither intrinsic nor representational in the ordinary sense of such terms. Let me elaborate.

First, an intrinsic property of a concept or thought is here understood as any property that does not depend on relations to other concepts or thoughts: “intrinsic representational features are those which can be stated without reference to another representation” (Gray 2017, 4). A concept token or thought token being of a certain type is thus to be understood as an intrinsic feature.<sup>19</sup> This is the case even if the type is determined by historical facts or other relations that are not strictly speaking intrinsic to the concept or thought on an ordinary understanding of ‘intrinsic’. The relevant notion of ‘intrinsic feature’ includes all features a concept or thought has when seen in isolation from other concepts and thoughts.

Second, the properties in questions need not be representational in the ordinary sense. If, for instance, one holds that mental representations are type individuated in terms of their historical origins<sup>20</sup> this would count as a feature of the relevant kind even though having a certain historical origin would not count as a representational feature on an ordinary understanding of ‘representational’. Likewise, if mental representations are individuated in terms of their syntactic *shape*<sup>21</sup> this would also be an intrinsic representational feature in this context. In general, the relevant features need not have anything to do with the fact that a thought represents an object. To repeat: The intrinsic representational features in question are those that can be characterized without reference to other mental representations.

---

<sup>19</sup> Unless the type is determined holistically.

<sup>20</sup> This is what Originalists claim. I say more about this view in section 3.4. below.

<sup>21</sup> I say more about this in the next sub-section, where I present Fodor’s LOTH.

We may thus define Intrinsicism about coordination thus:

**Intrinsicism:** Coordination is to be accounted for in terms of intrinsic representational properties of concepts and thoughts,

where the relevant properties are those that a concept or a thought has independently of other concepts or thoughts. This has been the dominant view of coordination until recently.

On the other hand, there are those who hold that coordination cannot be accounted for purely in terms of intrinsic representational features. According to such views, coordination is essentially a matter of relational representational features. Whether or not two thoughts are positively coordinated depends on how the thoughts are related, and such relations cannot be reduced to sameness or difference of intrinsic representational features.

We may thus define Relationism about coordination thus:

**Relationism:** Coordination is to be accounted for in terms of primitive relational representational features.

My positive account is a version of relationism. Even so, it bears some key similarities to certain well-known intrinsicist accounts. I now turn to an overview of the competing frameworks that I will address in this thesis.

### *3.2. Fodor's Language of Thought Hypothesis*

The key claim of Fodor's LOTH is that thinking takes place in a mental language. This language is distinct from natural languages, but bears some important similarities to such languages. The main similarity is that the language of thought also has a combinatorial syntax and semantics. A thought is made up of simple expressions that combine into complex sentence-like structures. The semantic content of a thought depends on the semantic content of the syntactic constituents of that thought:



Mental representations [...] have a *combinatorial syntax and semantics*, in which (a) there is a distinction between structurally atomic and structurally molecular representations; (b) structurally molecular representations have syntactic constituents that are themselves either structurally molecular or structurally atomic; and (c) the semantic content of a (molecular) representation is a function of the semantic contents of its syntactic parts together with its constituent structure. (Fodor & Pylyshyn 1988, 12)

The syntactic features of a thought are physically implemented in the brain. The cognitive system is only directly sensitive to such syntactic features of thoughts rather than their semantic properties. On this picture, the cognitive system resembles a computer in the sense that the cognitive system computes on syntactically specified entities. Fodor says that

Computations are operations defined over *syntax* of mental representations; it is the syntax, rather than the content, of mental states that determines its causal powers. (Fodor 2008, 70)

The LOTH thus implies a certain view of the metaphysics of thoughts. Propositional attitudes do not involve thinkers standing in direct relations to the relevant propositions. Instead, thinkers only stand in direct relation to mental representations having or expressing such propositional content. The LOTH can be understood as a conjunction of the following three theses:<sup>22</sup>

- A) **The Representational Theory of mind (RTM):** Thinking consists of causal sequences of tokenings of mental representations.
- B) **The Sentence-Like Structure of Thoughts:** Mental representations have a combinatorial syntax and semantics. Thoughts are built up of atomic constituents and the semantic content of a thought is a function of the semantic content of the constituent atomic representations and the syntactic structure of the thought.

---

<sup>22</sup> C.f. Lycan & Prinz (2008, 146).

- C) **The Computational Theory of Mind:** The cognitive system is similar to a computer in that it computes over syntactically specified mental representations.

Thesis (A) suggests a different picture of the nature of thoughts than the one we found in the previous section when discussing Frege and Russell. According to this general view, which I called the Dyadic View, thinking is a direct relation between a subject and a proposition. In contrast, according to RTM thinking is a triadic relation between a subject, a mental representation and propositional content. This has important consequences for how we account for Frege's puzzle. If we can account for the cognitive role of thoughts and concepts in terms of mental representations, we need not postulate a further level of semantics akin to Fregean senses. Nor do we have to follow Russell in his restrictivism about singular thoughts.

Remember that the doctrine of misidentification is what led Frege and Russell to develop their specific views on the nature of thoughts. According to this doctrine, it is the case that, if it is rationally possible for a thinker to take two thoughts to concern distinct individuals when they do in fact concern the same individual, there must be a difference in the two thoughts. On the RTM framework, it is possible to do justice to this doctrine without necessitating a difference in the propositional content of the relevant thoughts. Since subjects are only directly related to mental representations it is rationally permissible to be wrong about co-reference as long as the representational constituents of the two thoughts differ. If someone has the beliefs BOB DYLAN IS A MUSICIAN and ROBERT ZIMMERMAN IS A MUSICIAN she may, as we have seen, rationally fail to know that the two beliefs concern the same individual. However, all that follows directly from this on the current account is that the mental representations, i.e. the syntactic features of the thoughts, must be different. It does not follow directly that the two thought must differ in their propositional content.

This means that according to Fodor's LOTH, coordination is ultimately explained in terms of a sameness or difference in the syntactic features of thoughts. Whether or not two thoughts are positively coordinated depends on the

sameness or difference in the constituent syntactic entities of such thoughts. Such sameness or difference is, according to Fodor, a matter of two atomic representation tokens being of the same syntactic type.

The question, then, is how to individuate such atomic mental representations. Fodor is not explicit about exactly how to do this, but in some places he suggests that they are to be individuated by their syntactic *forms* (cf. Fodor 2008). Talk of such forms is of course purely metaphorical. Type identity of atomic representations is not determined by their physical 'shape' or 'form' in the brain, since the same concept may be physically realized in many different ways in the heads of different thinkers, or even the same thinker at different times. Fodor (2008) illustrates sameness or difference in atomic representation types by appeal to how we determine sameness of types of letters:

We distinguish 'dog' tokens from 'cat' tokens by their spelling, but we don't distinguish 'a' tokens from 'b' tokens that way, since 'a' and 'b' don't, of course, *have* spellings. What they have is *shapes*; and their shapes are different in ways to which our visual system is responsive; if they weren't, we wouldn't be able to read. Likewise *mutatis mutandis* for the way the minds draw type distinctions between tokens of basic mental representations. (Fodor 2008, 80)<sup>23</sup>

Exactly what is meant by the talk of shapes of syntactic entities is not important for the discussion in this thesis. Rather, what is important is that Fodor's account of coordination explains the phenomenon in terms of intrinsic representational features of mental representations, as defined in the previous section. Such intrinsic representational features, recall, are those features that are independent of the properties of other mental representations of a thinker. On this view, a mental representation being of a certain type is independent of its relation to other thoughts or concepts the thinker may have. Fodor's LOTH is thus an intrinsicalist account of coordination in thought.

---

<sup>23</sup> Fodor also uses this way of talking about mental representations in earlier work, such as in the following passage: "Because Classical mental *representations* have combinatorial structure, it is possible for Classical mental *operations* to apply to them by reference to their form" (Fodor & Pylyshyn 1988, 13).

The next two frameworks I will present, namely the mental file account and Originalism, also accept RTM. They hold that thinking is not a direct relation between subjects and propositions. The frameworks differ, however, with respect to their views as to the exact nature of coordination. They take different views on the nature of mental representations. I turn presently to the mental file account.

### 3.3. *Mental Files*

The mental file account of the nature of thoughts has been developed in various ways by different authors. One of the central proponents of this framework is Perry (1980). Fodor also mentions the possibility of modelling mental representations as files (2008, 92—100). The details of the framework are, however, most thoroughly developed by Recanati (2012, 2016).<sup>24</sup> In what follows, I will thus focus on Recanati's particular formulation of the framework.

According to this framework, singular concepts are to be understood as clusters of information construed as mental files. Each file contains pieces of information that are taken by the subject to concern the reference of the file. On this framework, coordination is explained in terms of such information clustering within mental files: All pieces of information contained within a single file are positively coordinated.<sup>25</sup> What file a given piece of information is stored in wholly depends on the acquaintance relation through which it was gained (Recanati (2012) uses the term *epistemically rewarding (ER) relation* for such relations). All pieces of information gained through the same acquaintance relation go into the same file and thereby become positively coordinated.

---

<sup>24</sup> Talk of files and equivalent notions is also used in discourse representation theory (e.g. Kamp 2015) and file change semantics (c.f. Heim 1982).

<sup>25</sup> Note that Recanati does not use the term 'positive coordination'. Instead he uses what I take to be a synonymous term, namely '*de jure* co-reference'. I say more about this terminology in section 6.1 below.

The mental file theorist maintains that such acquaintance (i.e. ER) relations may obtain between a thinker and objects in the external world.<sup>26</sup> Hence, the mental file framework differs from that of Russell in that one may have singular thoughts about mind-external objects. In fact, Recanati develops his account of mental files specifically as a way of accounting for the Fregean data within a Millian/Russellian framework.<sup>27</sup> He argues that we may have singular thoughts about mind-external objects and that this involves deploying a singular file in thought.

A mental file contains information in the form of predicates taken by the subject to be satisfied by the referent of a file. So, for instance, your mental file about Bob Dylan contains information such as ‘is a musician’, ‘won a Nobel Prize’, and so on. On this picture, two pieces of information, *i* and *j*, are positively coordinated if “*i* and *j* occur in the same file without the benefit of a prior judgement of identity” (Recanati 2012, 95). Every piece of information that is taken to concern the same referent is stored within the same file. This means that the fact that the two pieces of information ‘is a musician’ and ‘won a Nobel Prize’ is stored within one and the same file is what warrants an inference directly from the two pieces of information to the conclusion that a musician won a Nobel Prize. That is, co-location of information in files warrants trading on identity.

In contrast, two pieces of information being stored in distinct files is an indication that the information is taken to concern two different referents. Let’s say that instead of having both ‘is a musician’ and ‘won a Nobel Prize’ stored in the same file, you have the first predicate stored in your BOB DYLAN file and the

---

<sup>26</sup> Such ER relations may be perceptually based, but can also obtain in cases in which one is indirectly related to an object through testimony.

<sup>27</sup> Note that Recanati takes his mental file framework to be ‘neo-Fregean’. He says that he assumes “a two-level semantics with a sense-reference distinction” (2012, 13). However, in a footnote on the same page he says that “since, in my framework, mental files are what plays the role of sense, and mental files are representational ‘vehicles’, it can be argued that the theory put forth in this book is *not* a two-level semantic à la Frege” (2012, 13 n.6). Later on he says that “some theorists, most prominently Jerry Fodor, reject the Fregean distinction between sense and reference on the grounds that what plays the mode of presentation role is not anything semantic [...] So what is the difference between the view I have expounded and the view, argued for by Fodor [...] that modes of presentation are syntactic? Not much, since I accept that mental files are representational vehicles. The difference is primarily terminological” (2012, 244–245). For the purposes of this thesis, I thus take Recanati’s framework to be what I’ve called ‘Millian/Russellian’ in nature. Nothing hangs on this terminological difference.

second predicate stored in another file, namely your ROBERT ZIMMERMAN file. Let's also assume that you are not aware that Dylan is Zimmerman. Now, since the two pieces of information are stored in distinct files, you are not warranted in inferring from these two pieces of information alone that a musician won a Nobel Prize. The information being stored in distinct files entails that the two pieces of information are not accessible from one and the same file: "to say that there are two distinct mental files is to say that the information in one file is insulated from information in the other" (Recanati 2012, 42). This is why one cannot, without violating rational norms, trade upon identity of two pieces of information that are stored in distinct files.

This is also what accounts for the possibility of rational individuals ascribing contradictory predicates to the same individual: Since a difference in files makes it the case that the information is insulated, there is nothing that prevents you from storing the predicate 'is a musician' in the BOB DYLAN file and 'is not a musician' in the ZIMMERMAN file, since the contradiction is not manifest to you. On this picture, "information integration and inferential exploitation of information only takes place within files" (Recanati 2012, 43). Negative coordination, then, is explained in terms of information being stored in distinct files.

Note that information being gained through a given acquaintance relation counts as an intrinsic representational feature. This is because being gained on the basis of a given acquaintance relation does not depend on relations to other representations. Likewise, being stored in a given file does not depend on relations to the other pieces of information that may be in that file. Hence, the mental file framework also offers an intrinsicist account of coordination in thought. This feature of the mental file framework will play an important role in the discussion of this framework in this thesis.

I now turn to a discussion of Originalism, a view that, like the mental file account and Fodor's LOTH, holds that thinking does not involve a direct relation between subjects and propositions, and that coordination is to be accounted for in terms

of intrinsic representational properties of concepts and thoughts. However, it disagrees with these other accounts about the nature of mental representations.

### *3.4. Originalism*

Originalism is developed and defended by Sainsbury & Tye (2011, 2012). The framework is specifically developed to account for the Fregean data within a Millian framework. Sainsbury & Tye explicitly appeal to possible worlds semantics when illustrating the nature of the content of thoughts (2012, 47—46). However, the general framework would be equally compatible with classic Millian/Russellianism about propositions. What's relevant in the current context is that they claim that mental representations can do the work traditionally assigned to Fregean senses, and so there is no need to introduce further levels of semantics in order to account for coordination.

Originalists accept certain key claims of Fodor's LOTH. First, concepts are taken to be mental representations. Second, Originalists agree with Fodor that thoughts have a combinatorial structure that consists of atomic concepts and relations between such concepts:

[Concepts are] mental representations of a sort deployed in thought; they are representational constituents of thoughts. Thoughts are made up of concepts, and what thoughts as a whole represent is a function of their component concepts: what they represent and how they are combined. (Sainsbury & Tye 2012, 1)

Originalists also agree with Fodor that the cognitive system computes over token mental representations: "Cognitive processing depends not directly on content but on the vehicles of content: concepts and thoughts" (Ibid., 57). Hence, Originalists take thinking to be a triadic relation between a thinker, mental representations and propositional content. Thinkers are directly related to mental representations and only indirectly to the propositional content of such representations.

The view differs, however, from that of Fodor in what it takes the individuation conditions for concepts to be. The key claim of Originalism is that two concept tokens are of the same type if and only if they have the same historical origin:

**Originalism:** Concept C1 = concept C2 if and only if the originating use of C1 = the originating use of C2 (c.f. Sainsbury & Tye 2011, 105).

On this view, it is the case that for every concept there is just one originating use and that every originating use of a concept is the origin of one concept only (c.f. Sainsbury & Tye 2011, 104). So for instance, in the case of someone tokening the concept ROBERT ZIMMERMAN, the type of the concept is determined by the time in history at which it was first introduced. In this case, this would most likely be at Bob Dylan's baptism. Similarly, the type of a given tokening of BOB DYLAN is determined by the first use of that term, which took place in a different historical context than the first use of ROBERT ZIMMERMAN. As a result, the two concept tokens in question are of different types, due to having distinct historical origins.

On this view, coordination is explained in terms of sameness or difference in concept types. Concept tokens that are of the same type play the same role in cognition, while concept tokens of different types play distinct roles in cognition. Whether or not two concept tokens are of the same type fully depends on whether or not they belong to chains of deference with the same historical origin. Importantly, Originalists thus reject Frege's (1982) claim that the possible cognitive difference of co-referential concepts requires the introduction of senses. They hold that,

distinct concepts can, and typically will, play different roles in our cognitive activities, even if they have the same content. [...] The work supposedly done by difference of sense can be done better by difference of concepts. (Sainsbury & Tye 2012, 53—54)

This is the same general idea that we found in Fodor's LOTH and the mental file framework: We may adopt a simple one-levelled semantics and still account for the Fregean data by appeal to sameness or difference in types of mental representations.



Since Originalists explain coordination in terms of sameness or difference in types of concepts (and such types are not determined holistically) they ultimately account for coordination in terms of intrinsic representational features of concepts and thoughts, as defined in section 3.1. To determine the type of a concept one need not look to other concepts the thinker deploys. We may determine the type of a concept purely by looking at historical facts about when the concept was first used. In the next section I will say more about why this point is important to the discussion in this thesis.

### *3.5. Semantic Relationism*

As mentioned throughout the presentation of the three previous views, they all have in common that they explain coordination in terms of intrinsic representational features of concepts and thoughts. Such intrinsic representational features, recall, are those that can be established by looking at a particular mental representation in isolation from other mental representations. According to this terminology, then, a mental representation having a certain content (i.e. referent) or being of a certain type (as long as types are not determined holistically) counts as an intrinsic representational feature of that representation (and, as we have seen, this is the case even when type individuation involves an appeal to historical fact).

In recent years, some philosophers have argued that this approach to coordination is misguided. Such philosophers (c.f. Fine 2007, Pinillos 2011) argue that coordination cannot be reduced to sameness of intrinsic representational features. This is because they take coordination to be an essentially relational property. I call this general account of coordination in terms of irreducible primitive relations *Relationism*. The most prominent relationist account is found in Fine (2007), where he develops and defends his *Semantic Relationism*. Fine develops relationist accounts for coordination in

logic, natural language and in thought. In this thesis I focus on the latter, namely coordination in thought.<sup>28</sup>

Like the views discussed in the previous subsection, Fine's account is advanced as a way of accounting for coordination within a broadly Millian/Russellian framework. It differs, however, in how the framework achieves this. Fine does not ascribe any interesting explanatory role to mental representations when it comes to coordination in thought. This is partly because he thinks that "thoughts do not appear to have the same kind of clear syntax as sentences" (2007, 73). This marks a clear discontinuity with the previous views which all take thoughts to consist of structured representations and explain coordination in terms of features of such mental representations. When it comes to Fine's account of coordination we may thus regard his view as closer to the Dyadic View in that, at least for our explanatory purposes, thinking is a dyadic relation between a subject and a proposition.

Consequently, Fine follows Frege and Russell in taking a difference in cognitive role to imply a difference in propositional content. Importantly, however, Fine takes the relevant difference needed to account for coordination to be a primitive relational feature of propositions. Two thoughts that differ in their cognitive role may share all intrinsic representational features and differ only in their relational properties. On this view,

The content of a belief will be given by a coordinated rather than by an uncoordinated proposition. Thus we may distinguish between the content of the belief that Cicero is Tully (where this is the negatively coordinated proposition) from the content of the belief that Cicero is Cicero (where this is the positively coordinated proposition). This is already a great advantage on the usual referentialist view, which is

---

<sup>28</sup> As I read Fine, he thinks that it is possible to consider coordination in the different domains in isolation. Even though Fine's explanation of coordination is highly similar across these domains, what's important for the present purposes is that his explanation of coordination in thought does not depend on what he says about the other domains. That is, although his explanations of coordination in language and thought are similar, his account of coordination in thought does not essentially depend on his account of coordination in language and vice versa. I say more about the relation between coordination in thought and language in section 7.4. below.

unable to make any such distinction without either distorting the logical form or appealing to some notion of sense or “guises”. (Fine 2007, 77)

Even though two thoughts such as DYLAN IS DYLAN and DYLAN IS ZIMMERMAN express the same *classic* Millian/Russellian propositions, they express different *coordinated* propositions: The former proposition contains a coordinative link between the two objectual constituents making it positively coordinated, while the latter proposition does not contain such a link, making it negatively coordinated. This is why in the former thought the co-reference is manifest to the thinker while in the second thought this is not the case.

An important consequence of Semantic Relationism, then, is that even though classic Millian/Russellian propositions are not transparent to thinkers, coordinated propositions must be. I return to this in section 6.2 below.

The account of coordination developed in this thesis bears many similarities to the views discussed in this general section. In the next section I will give a brief overview of key-claims of the proposed account and point to similarities and differences between this framework and the related frameworks considered above.

#### 4. Vehicle Relationism

I develop the framework of Vehicle Relationism throughout chapters 1, 2 and 3. In what follows I will point to some of the central features of the framework.

The key claim of Vehicle Relationism is this:

**Vehicle relationism:** Coordination is to be accounted for in terms of primitive relations between representational vehicles.

I use the notion of *pointer relations* to denote the relevant primitive relations. Whenever the pointer relation obtains the relata are positively coordinated.

Importantly, the pointer relations do not reduce to sameness of intrinsic representational features of the relata.

I call the representational vehicles that are capable of entering into pointer relations *mental tags*. Together with pointer relations, mental tags are the building blocks of thoughts, and as such they can be understood as individual concepts.

We then get the following characterization of coordination relations:

**Positive Coordination:** Two mental tags are positively coordinated if and only if they are connected by pointers.

**Negative Coordination:** Two mental tags are negatively coordinated if and only if they are not connected by pointers.

Vehicle Relationism thus assumes a representational theory of mind. According to this view, thinking is a triadic relation between a subject, mental representations, and propositional content. Thinkers are only directly related to mental representations and indirectly to propositions via such mental representations. The cognitive computational system is only directly sensitive to features of mental representations. In this way, the framework bears similarity to that of Fodor's LOTH, the Mental File Framework, and Originalism.

It differs from such frameworks in that it is essentially a relationist account of coordination. According to Vehicle Relationism, it is not possible to determine whether or not two representational vehicles are positively coordinated merely by looking at the properties of the two vehicles in isolation. Two representational vehicles may share the exact same intrinsic properties – including the semantic content they express – and still fail to be positively coordinated. In the case of someone rationally having both the belief DYLAN WON A NOBEL PRIZE and the belief ZIMMERMAN DID NOT WIN A NOBEL PRIZE, the mental tags DYLAN and ZIMMERMAN have the same intrinsic properties, including referential content, but they are not connected by pointers. Since they are not connected by pointers, the co-reference is not manifest to the thinker.

Vehicle Relationism thus agrees with Semantic Relationism in that coordination is to be accounted for in terms of primitive relations. The framework differs from that of Fine's in that it takes the relata of these relations to be representational vehicles, whereas Fine takes the relata to be objectual constituents of propositions.

Since the framework strictly speaking is an account of mental representations it is in theory compatible with any account of the nature of propositions. However, as already mentioned, I accept a Millian/Russellian account of propositions, and accounting for the Fregean data within a Millian/Russellian framework is one of the central motivations behind the development of Vehicle Relationism. The viability of Vehicle Relationism does not, however, depend on the viability of this particular view on the nature of propositions.

Thus far, I've focused on a notion of samethinking essentially tied to the explanation of rational reasoning and behavioural dispositions. In particular, I have focused on coordination, which is to be understood as the sort of samethinking that essentially has to do with our capacity to trade on identity. In the final chapter in this thesis I explore a different way in which thoughts can be said to be about the same phenomenon. This notion of 'samethinking' is tied in with a related notion of 'samesaying'. Since the latter has received much more attention in the literature, I focus on utterances rather than thoughts when discussing this broader notion. However, as far as I can see, most of what I say is equally applicable to the phenomenon as it occurs in thoughts. The question I address in the final chapter is how someone (typically distinct individuals or the same individual at different times) may properly be said to think or talk about the same topic despite their thoughts and utterances differing in their overall semantic properties. In the next section I will give some more background to the debate on topic-continuity in cases of semantic drift.

## 5. Topics

The final chapter of this thesis concerns the broader understanding of ‘samethinking’ as well as the corresponding notion of ‘samesaying’. It focuses on the possibility that two thoughts or two utterances may concern the same topic despite a difference in their overall semantic properties. The discussion in this final paper does not build on the discussion in the previous chapters. It does, however, concern the same general abstract theme: What is it for two thoughts/utterances to be ‘the same’ in explanatorily interesting ways?

The chapter engages with the recent literature on Conceptual Engineering.<sup>29</sup> Conceptual Engineering is the process of assessing and improving our representational devices. In the chapter, I consider prominent questions within this literature and engage with some of the central views on the nature of the project. I do, however, combine this literature with considerations familiar from the literature on the nature of thoughts and language.

Semantic drift occurs whenever a word changes meaning over time. This is a common phenomenon. Consider for instance the word ‘clue’, which used to mean a ball of thread (of the sort that would help guide someone out of a labyrinth, for instance), whereas today it means something like evidence or information that helps solve a given task. It seems that the meaning of the word has changed. This kind of semantic drift often results in a change in the subject matter that one talks (or thinks) about when deploying such words. This is the case with the word ‘clue’: People who used the word at a time when the word picked out balls of threads specifically did not talk about the same subject matter as someone who uses the word today. In cases like this, there has been a change in topic as a result of the semantic drift.

Some cases of semantic drift, however, preserve topic. That is, there are cases in which it seems correct to say that a use of a word prior to a semantic change concerns the same subject matter as a use of the very same word type after such

---

<sup>29</sup> C.f. Haslanger 2000, Plunkett & Sundell 2013, Cappelen 2018. Plunkett and Sundell (2013) use the label ‘conceptual ethics’, while Haslanger (2000) uses the term ‘ameliorative project’.

a change. At the beginning of this introduction chapter, we looked at one such candidate. In the case of the word ‘fish’ it seems right to say that someone who used the word at a time when they thought of whales as a species of fish talked about the same subject matter as someone who uses the word today, although such a person would not consider whales a type of fish. This is the case even if we think that the meaning of ‘fish’ has changed as a result of people coming to think that whales are not fish.<sup>30</sup>

In recent years the notions of semantic change and stability of topic have gained currency as a result of the increased popularity of revisionary projects such as conceptual engineering. Sometimes our representational devices do not work the way we want them to.<sup>31</sup> There are various ways in which such devices may be sub-optimal: For instance, our words may fail to refer; or they may refer to more things than we would like them to; or they may fail to pick out things that should be in their extension; they may have negative consequences for a society and so on. Philosophers are not merely in the business of describing the nature of our representational devices – sometimes they ask what our representational devices *should* be like.

Revisionary projects, such as conceptual engineering, involve changing semantic aspects of representational devices. There is, however, a worry that such changes will lead to a change in subject matter, resulting in people talking past each other and providing solutions to problems other than the ones originally posed. This is the Strawsonian challenge (c.f. Strawson 1963). In order to respond to this challenge, many philosophers nowadays (e.g. Sawyer 2018, Cappelen 2018) think that there may be a continuity of topic despite a change in semantic properties. In the last chapter of this thesis, I consider some prominent theories of how to best account for the possibility of continuance of topics despite such semantic drift. I will say more about this discussion as well as my argument in the next section, where I give an overview of each of the thesis chapters.

---

<sup>30</sup> I borrow the case of ‘fish’ from Sainsbury (2014).

<sup>31</sup> I use the term ‘representational device’ rather than ‘concept’ since the former is less committing and theory-laden.

## 6. Chapter Overview and Summaries

The thesis consists of four papers. The first three papers concern coordination in thought. The fourth paper concerns the nature of the continuance-of-topic relation. In what follows I will give a brief overview of the discussion and argument of each chapter in turn.

### *6.1. Chapter 1: 'Mental Files: In Defence of Pointer Relations'*

The first chapter proposes a novel way of understanding coordination. It takes the mental file picture (c.f. Recanati 2012, 2016) as an outset and develops a more parsimonious framework for accounting for the nature of coordination in thought. I argue that the mental file framework is insufficient when it comes to accounting for coordination of relational predicates. In the case of relational predicates coordination relations obtain across distinct files and the mental file theorist cannot account for this purely in terms of co-location of information in a single file.

I consider various responses on behalf of the mental file theorist and argue that in order to give an account of coordination of relational predicates, the file framework must be supplemented with further machinery. More specifically, I argue that the mental file theorist is forced to introduce *pointer relations* in order to account for this phenomenon. Such pointer relations account for the possibility of positive coordination across distinct files.

I then argue that this further machinery, i.e. the pointers, in effect renders the files themselves superfluous when it comes to accounting for coordination in general. If pointers can account for coordination across distinct files they can also account for coordination within files. If we grant this, the files themselves are no longer needed to account for coordination of one-place predicates. We may give a unified account of coordination in terms of pointer relations without



invoking the notion of a mental file. I conclude that we ought to give up the explanation of coordination in terms of co-location of information in files and instead adopt a relational account of coordination in terms of pointer relations.

I call the proposed framework *Vehicle Relationism* (c.f. section 4 above). This framework has many common features with the mental file framework and is developed partly as a result of engagement with the literature on mental files. The main difference between the two frameworks is that the mental file framework is an intrinsicist framework while Vehicle Relationism is a relational framework. The mental file theorist explains coordination in terms of sameness or difference in files, whereas Vehicle Relationism takes the pointer relations as representational primitives that cannot be reduced to sameness of intrinsic features of mental representations.

In chapter 1 I spell out the key claims of Vehicle Relationism. However, the view is only worked out in detail in chapter 2. Before giving an overview of chapter 2, let me make a brief terminological point. When discussing coordination in Chapter 1 I use the term '*de jure* co-reference' rather than 'positive coordination'. The terms pick out the same phenomenon, namely the relation that warrants trading on identity.<sup>32</sup> The reason why I choose to use this term in the discussion of the mental file framework is that this is the term used by Recanati himself. This makes the discussion and references to the mental file literature flow more easily. I ask the reader to keep in mind that there is only a terminological difference (hence, no substantial difference) between *de jure* co-reference and *positive coordination*.

## 6.2. Chapter 2: '*Vehicle Relationism: In Defence of Pointer Relations*'

The second chapter develops the framework suggested in Chapter 1 further. Here I spell out the details concerning the cognitive role of mental representations on this framework. The key claims of Vehicle Relationism are

---

<sup>32</sup> Recanati also notes this synonymy when he says that, "coordination' is another name for co-reference *de jure*" (Recanati 2016, 33).

the following. First, coordination is to be accounted for in terms of primitive relational features of thoughts. Second, such relations obtain between representational vehicles. I develop the details of Vehicle Relationism by appeal to the notion of pointer relations discussed in Chapter 1.

In this chapter, I use the positive view to account for certain long-standing puzzles within the philosophy of mind, such as Frege's Puzzle as well as Kripke's puzzle of Paderewski. I argue that in order to solve these puzzles we need only appeal to relational differences of the relevant mental representations.

I then compare the positive view to similar views, such as Heck's (2012) *Formal Relationism*, Fodor's (1975, 2008) LOTH, and Fine's (2007) Semantic Relationism. The key claim of Heck's formal relationism is that all that is needed to account for the Fregean data is relational differences between the relevant thoughts. Heck does not say much about how to implement such relations. They do, however, say that they have an inclination to think that "the language of thought hypothesis is true and that formal relations supervene on Mentalese syntax" (Heck 2012, 159). While I agree with Heck's general observation that relational aspects of thought are sufficient to explain the Fregean data, I disagree with their view about how to implement such relations if the syntax of the language of thought is construed according to Fodor's intrinsicist view. This leads to a comparison between Fodor's LOTH and Vehicle Relationism.

As we have seen, Fodor's LOTH and Vehicle Relationism agree on several central issues. The key difference concerns how the two frameworks understand the language of thought. Fodor takes the elements of the language of thought to be highly similar to sentences and words in natural language and thus takes coordination to be explained in terms of sameness of types (which is determined by something akin to shapes) of the "words" in the language of thought. In contrast, the Vehicle Relationist takes the relational aspect of mental representations to be representationally primitive and an essential constituent of the language of thought. On this view, the cognitive system "reads off" the pointer relations directly, rather than recognizing coordination as a result of encountering the same or different symbols in the language of thought.

Finally, I compare Vehicle Relationism to Semantic Relationism. I point to problems with Semantic Relationism that are easily avoided by the proposed account. In short, I argue that it is puzzling how coordinated propositions can be transparent to thinkers whereas classical Millian/Russellian propositions are not. I also argue that Vehicle Relationism is more metaphysically parsimonious than Semantic Relationism and that, as a result, Vehicle Relationism provides a better picture of coordination in thought.

6.3. *Chapter 3: 'Originalism and Coordination in Thought: In Defence of Vehicle Relationism'*

Chapter 3 is to be seen in line with the former two chapters. In this chapter I offer criticisms of one particular account of coordination, namely Originalism (c.f. section 3.3. above). I argue that the Originalist fails to account for coordination in thought and that we instead should adopt a Vehicle Relationist account of such coordination.

I consider two constraints on concept individuation that are generally accepted in the literature, but that are jointly inconsistent. First, according to the *Publicity Constraint* on concept individuation, the nature of concepts must be such that distinct individuals may use the same concept or concept tokens of the same type. This constraint calls for a coarse-grained individuation of concepts.

The second constraint is the *Fregean Constraint*. According to this constraint, a difference in the cognitive role of thoughts and concepts must be reflected in a difference in concepts or concept types (c.f. section 2 above). This constraint requires fine-grained individuation conditions for concepts. The two constraints are thus in tension (c.f. Crimmins 1992, Heck 2002, Laurence & Margolis 2007, Duahu 2012, Onofri 2016).

Given the nature of Originalism, the framework seems *prima facie* promising when it comes to accounting for the two constraints. First, the framework takes concepts (types) to be public and hence shareable. This is promising when it

comes to accounting for the Publicity Constraint. Second, the framework is specifically branded as providing solutions to the longstanding puzzles pertaining to the cognitive role of concepts and thoughts. It should therefore be able to account for the Fregean data.

I argue, however, that Originalism fails to provide a good account of the cognitive role of concepts and thoughts. The problem is that the Originalist is forced to say that sameness or difference in concept type is not always transparent to thinkers. Coordination relations are, however, always transparent. Hence, Originalism fails to provide an account of coordination and thus to give an account of rational reasoning. As a result, the framework fails to provide satisfying solutions to classical puzzles of mind, such as Frege's Puzzle and Kripke's (1979) puzzle of Paderewski.

The problem posed for Originalism is a structural problem that affects any theory that takes the cognitive role of concepts to be accounted for in terms of types of concepts, but at the same time holds that sameness or difference in concept types is not transparent to thinkers. I suggest a minimal addition to such frameworks that allows them to account for a thinker's rational cognitive capacities. I propose that the cognitive role of concepts is to be accounted for in terms of primitive relations (i.e. pointer relations) that obtain between representational vehicles rather than in terms of features such as representations being of the same type. This opens up the possibility of having coarse-grained individuation conditions for concepts (and thus respect the Publicity Constraint on concept individuation), and at the same time do justice to Frege's observation without strictly speaking conforming to the Fregean Constraint.

#### *6.4. Chapter 4: 'Staying on Topic: The Continuance-of-Topic Relation is Non-Transitive'*

In chapter 4 I look at the phenomenon of continuance of topics in cases of semantic drift. There has been a tendency to make a sharp distinction between,

on the one hand, cases of semantic drift where the topic is preserved and, on the other, cases of semantic drift where there is a change in topics (e.g. Sawyer 2018, Ball forthcoming). When it comes to accounting for continuance of topic the dominant strategy has been to appeal to identity of some sort. I show that we cannot draw a clear distinction between the cases of semantic drift where topics are preserved and those where they are not. This is because we have cases that in some sense fall into both categories. There are cases of semantic drift where there is continuity of topic at any two minuscule time intervals through the evolutionary chain of a term, but where there is discontinuity of topic at the beginning and end of the chain. That is, the continuance-of-topic relation is non-transitive.

By showing that the continuance-of-topic relation is non-transitive, I reveal a structural problem with popular accounts of topic stability according to which continuance of topic is accounted for in terms of identity relations, be it identity of concepts (Sawyer 2018, Richard forthcoming) or identity of meaning (Ball forthcoming). I argue that all accounts that try to explain stability of topics in terms of identity relations fail.

Finally, I consider Cappelen's (2018) account of stability of topics, and argue that the non-transitivity of the sameness-of-topic relation puts some restrictions on his *Contestation Theory*. I argue that a consequence of the finding in this paper is that proponents of Cappelen's *Contestation Theory* can only account for continuance of topics if they adopt a similarity account of samesaying. They are thus forced to reject Cappelen & Lepore's (2007) argument to the effect that the samesaying locution is necessarily a matter of identity.

In the final section of this introduction chapter I will point to some of the findings in this thesis. I will also suggest some future work based on these findings.

## 7. Findings and Further Work

The thesis proposes a novel account of coordination in thought. I present the key claims of Vehicle Relationism and use it to explain some of the central puzzles pertaining to coordination within the philosophy of mind. I compare the view to what I take to be the main competing frameworks given certain common assumptions about the nature of thoughts. In what follows I will point to some of the consequences of the proposed framework as well as suggest some further work.

### *7.1. Consequences for Trading on Identity*

One interesting consequence of Vehicle Relationism is that, if correct, the validity of an inference and rational reasoning come apart. In presenting the debate in this chapter I have explained the relevant notion of ‘manifestness’ in terms of its role in giving warrant for an individual’s trading on identity. However, a consequence of Vehicle Relationism is that one may be warranted in trading on identity even in cases where the relevant thoughts do not in fact concern the same individual as long as the cognitive system (mistakenly) treats the thoughts as if they did concern the same referent. In Chapter 2, I draw a distinction between manifest and apparent co-reference:

**Manifest Co-reference:** Two mental representations are manifestly co-referential if and only if they are positively coordinated and share reference.

**Apparent Co-reference:** Two mental representations are apparently co-referential if and only if they are positively coordinated and do not share reference.

Both manifest and apparent co-reference are sub-classes of positive coordination. In order to be warranted in trading on identity one only needs positive coordination. In the case of apparent co-reference, someone may be warranted in trading on identity without the inference being valid. The semantic

constraint commonly taken to be the mark of valid inferences says that, if the premises are true, then so necessarily is the conclusion. However, in the case of apparent co-reference it is possible that the premises are true, but that the conclusion is false.

I take this to be a virtue of the proposed framework. This helps us explain how someone may reason in a rational way in cases of confusion. To illustrate, suppose you see a snake on the ground and form the belief THAT SNAKE IS VENOMOUS. You look away and when you turn back you see what you take to be the same snake again and form the belief THAT SNAKE ENJOYS SUNBATHING. You would then be rational to conclude from this that there is a venomous snake that enjoys sunbathing. However, unbeknownst to you, the apparent snake you're looking at turns out to actually be two distinct snakes. The two representations are positively coordinated and so you are rational in trading on identity in the inference. However, since the two positively coordinated representations refer to distinct individuals, the argument is not valid. A consequence of this view is that positive coordination is not factive, in that two mental representations that are positively coordinated may not be co-referential.

This contrasts with other views on coordination, such as Fine's (2007) and Recanati's (2016). According to Fine, positive coordination entails co-reference: "coordination within thought is taken to be a form of strict co-representation, in analogy to [...] coordination within language as a form of strict coreference" (2007, 67), and "strict co-reference implies coreference" (Ibid., 45). In the case of the snake above, he would either have to claim that the relevant beliefs are not positively coordinated or that they actually co-refer.<sup>33</sup>

---

<sup>33</sup> The latter alternative seems to be what Fine would prefer. He might say that all three tokens of THAT SNAKE in the beliefs figuring as premises and the conclusion have the same referent, namely an amalgam of the two snakes: In cases of confusing two objects for one, "we have successful reference to some sort of amalgam of these objects" (Fine 2007, 126). Fine also introduces the notion of *putative* co-reference which corresponds to what is taken to be a matter of co-reference from the subject's point of view. This notion is not factive (c.f. Fine 2010b, 497). Putative coordination explains why an individual may trade on identity in cases of confusion. In such cases, the individual may trade on identity even though the relevant beliefs are not positively coordinated (since positive coordination is a matter of strict co-reference on his view). I do, however, think that in order for coordination to be the right notion to explain cognitive significance in general (including cases of confusion such as in the case of inverse Paderewski puzzles (c.f. Recanati 2012, 115–116)) warrant in trading on identity and positive coordination should not come apart.

Similarly, Recanati takes positive coordination to be factive. He takes positive coordination to be a matter of knowledge such that, if two terms or token mental representations are positively coordinated the thinker *knows* that they co-refer if both refer.<sup>34</sup> He thus denies that cases of confusion, such as the one involving the snake above, involve reference to two distinct referents. Instead, he would say that only the first token of THAT SNAKE refers to the snake, whereas the other fails to refer. He says that “*if the subject is confused, at least one of the two singular terms must fail to refer*” (Recanati 2016, 28).

The Vehicle Relationist account of coordination thus differs from the accounts offered by Fine and Recanati in that the Vehicle Relationist does not take positive coordination to be factive in the sense that it yields (possibly conditional) knowledge of co-reference.

The fact that Vehicle Relationism allows for the possibility of positive coordination (and hence warrant in trading on identity) despite a difference in referential content makes the view similar to Lawlor’s 2010 account of *internal co-reference*.<sup>35</sup> According to Lawlor’s account “we might understand internal co-reference in terms of a chaining relation among token expressions in thought and language, with no implication of successful reference” (2010, 493). Lawlor’s account differs from that of Vehicle Relationism in that she takes such chains to obtain at the semantic level.

On the proposed framework, then, we only get validity in the case of manifest co-reference. Only in the case of manifest co-reference does the thinker potentially get new knowledge through an inference. But rationality may endure even in cases where this is not the case. On this picture, then, the notion of coordination in thought is first and foremost tied in with the rationality of an inference – the class of inferences that yield knowledge (in the relevant way) is only a subclass

---

<sup>34</sup> On this picture, positive coordination entails knowledge of *conditional* co-reference, i.e. that there is co-reference *if* the concepts refer at all (c.f. Recanati 2016, 21).

<sup>35</sup> See Lawlor (2010) for reasons to think that this is the correct understanding of positive coordination. She argues that we need “a substantive characterization that permits us to say how inferences might be warranted from the subject’s point of view even if they’re not truth-preserving [and] strict coreference doesn’t work for confused inference” (2010, 493). See Fine (2010b) for a response to Lawlor’s argument.



of inferences that involve positive coordination.

### *7.2. Consequences for the Notion of a Concept*

Another consequence of the view proposed and defended in this thesis is that talk about concepts as the constituents of thoughts is less explanatorily fruitful than often assumed. What would talk about concepts even amount to on the Vehicle Relationist account, one may ask. The question is a difficult one. One possibility is to take concepts to be mental tags. Mental tags are the bearers of semantic content, and the content of a thought depends on the constituent mental tags of that thought and how they are structured. If we take concepts to be tags of this kind, talk of concepts provides just another way to (indirectly) talk about the content of thoughts. The question, then, is what explanatory work this further notion of a concept might do that could not be done equally well by direct appeal to the content of the tags. Further, when it comes to the traditional job description usually assigned to concepts, such as accounting for cognitive significance, mental tags do not in and of themselves play an interesting explanatory role. Such explanatory tasks pertaining to coordination are all done by the primitive pointer relations. Another possibility, then, is to take concepts to be pointer relations. This, however, would lead to another non-traditional view of the nature of concepts. Concepts are usually taken to be the sort of things that can be given individuation conditions. However, when it comes to pointer relations, talk about sameness or difference in types is misguided. Pointer relations are specific instantiations of coordination relations in the mind of single individuals. There is not much sense to be made of the notion of two (token) pointer relations being of the same type.

The notion of a concept can, and has been, understood in a variety of different ways. Sometimes concepts are taken to be public and sharable and sometimes they are taken to be individual and hence non-sharable. By adopting Vehicle Relationism we may, so to speak, *free* the notion of 'concepts' usually taken to account for intrapersonal workings and reserve the notion for the explanatory

purposes pertaining to interpersonal phenomena. The result would be that there would no longer be a tension between the Publicity Constraint and the Fregean Constraint on concept individuation (c.f. Crimmins 1992, Heck 2002, Laurence & Margolis 2007, Duahu 2012, Onofri 2016), since the Fregean Constraint could then be abandoned. I say more about this in chapter 3 of this thesis.

In the remaining part of this chapter I will suggest further work prompted by the positive proposal in this thesis.

### *7.3. Further Work: Empirical Findings in Psychology*

In Chapter 1, I argue that explaining coordination in terms of pointer relations renders the mental file framework superfluous. The files are, however, designed to do more than just account for coordination. Since the central topic of the three first papers concerns the nature of coordination in thought I focus specifically on this explanatory task when discussing the mental file framework. However, in general, the frameworks are so similar that I see no reason to think that one cannot do with pointers what has previously been done with files. But this obviously requires discussion going beyond what I consider in the present thesis.

One use of mental files that I find particularly interesting is the deployment of the framework in explaining the cognitive development of young children in child psychology. In particular, psychologists have employed the notion of a mental file in attempting to account for young children's ability to pass false belief tasks (c.f. Perner et.al. 2015). Future development of Vehicle Relationism may involve an assessment as to whether the proposed framework can perform similar jobs. It would, in general, be interesting to see how well the framework fits with empirical findings in psychology. This is one direction which further research into the nature of pointer relation may take.

#### 7.4. Further Work: Coordination in Language

Another direction which future work on Vehicle Relationism may take is research into the possible implication of the framework for coordination in language. In this thesis I focus on coordination in thought rather than in natural languages. Given the close connection between thought and language, further research into the proposed framework might give new insights into coordination in language as well.

Here are some interesting questions concerning coordination that I will not consider in much detail in this thesis: What is the connection between coordination in thought and coordination in language? Do we need two separate frameworks in order to give an account of coordination in both realms? Can coordination in thought and language come apart in such a way that someone may utter a negatively coordinated sentence to express a positively coordinated thought and vice versa? If so, what is the coordinative status of the utterance as a whole?

Although the two domains are closely related, I take it to be unproblematic to propose an account of coordination in thought without reference to coordination in language in the way done in this thesis. It seems plausible that Recanati is right when he claims that “coreference *de jure*, even though it manifests itself in language, is first and foremost a phenomenon at the level of thought” (Recanati 2016, 10).<sup>36,37</sup> If this is right, it is likely that one may develop an account of coordination in thought without reference to coordination in natural language. However, it would seem to follow that the converse is not possible. According to the mental file framework, “coreference *de jure* at the language level is to be accounted for in terms of deployment of the same file in thought. The identity which grounds coreference *de jure* is not the identity of expressions but identity

---

<sup>36</sup> Again, *de jure* co-reference is the same as positive coordination.

<sup>37</sup> Burge (1979) argues that even Frege was primarily interested in thoughts when he developed his account of sense: “Frege was primarily interested in the eternal structure of thought, of cognitive contents, not in conventional linguistic meaning. He pursued this interest by investigating the structure of language, and much of his work may be seen as directly relevant to theories of linguistic meaning. But the epistemic orientation of his theorizing leads to a notion of sense with a different theoretical function from modern notions of meaning” (1979, 398-399).

of the mental file associated with them” (Recanati 2016, 12). So for instance, the positive coordination in the case of an anaphoric statement, such as “Lisa is home. She never left” is explained in terms of the same mental file being associated with both ‘Lisa’ and ‘She’. That, of course, is not to say that the identity at the conceptual level cannot be encoded by the syntax of natural language. Recanati suggests that there may be “*recurrence constraints on conceptual elements* [...] encoded in the syntax of natural language” (Ibid., 10).

A similar explanation is available to the Vehicle Relationist. On this view, one may hold that coordination in language is to be accounted for in terms of pointer relations between the mental representations associated with the relevant terms. So, for instance, in the case of a statement such as “Lisa is home. She never left”, ‘Lisa’ is associated with one particular mental tag instantiation and ‘She’ is associated with another instantiated mental tag, and these two mental tags are connected by a pointer. The pointer relation may be encoded by the syntax of the sentence, but ultimately the explanation of coordination of anaphoric statements depends on coordination in thought. Although this picture of coordination in language fits well with the proposed framework, the Vehicle Relationist is not committed to this particular view on coordination in language.

### *7.5. Further Work: Understanding the Exact Nature of the Pointer Relations*

I have said that the pointer relation is a primitive relation. By this, I mean that the pointers are *representationally* primitive. That is, they cannot be reduced to sameness or difference in intrinsic representational features of concepts and thoughts. The pointers are a representationally irreducible part of the language of thought.

That, of course, is not to say that pointer relations cannot be reduced to sameness or difference in *any* property. I have not said anything about how the pointer picture may be neurologically implemented, and make no claim about this matter in this thesis. It might very well be that the pointers ultimately obtain

## INTRODUCTION

or fail to obtain as a result of sameness or difference in neurological firings. I leave this as a topic for future work.

Although there is still more work to be done, I think the findings in this thesis provide a foundation for such future enquiry.

---

# De Jure Co-Reference: In Defence of Pointers

*According to the popular Mental File view, de jure co-reference is to be accounted for in terms of co-location of information within mental files. In this paper I argue that the mental file theorist faces problems in accounting for de jure co-reference of relational predicates in this way. I show that in the case of relational predicates she is forced to postulate a further notion – the notion of a pointer relation – to give an account of how de jure co-reference may obtain between pieces of information across distinct files. I then argue that it is this notion and not the notion of a file that accounts for de jure co-reference. The resulting view is a version of relationism, but it differs from other versions of relationism in that the relevant relations hold at the level of representational vehicles rather than at the level of content.*

## 1. Introduction

This paper offers an account of samethinking. As I use the term, samethinking occurs whenever two thoughts concern the same referent.<sup>1</sup> There are two

---

<sup>1</sup> By 'thoughts', I have in mind psychologically instantiated mental representations. It is natural to think I thus have in mind *thought tokens*. However, I avoid framing the problem in terms of thought types and tokens. There are two reasons for this: (i) it is controversial what the nature of

different ways in which this can occur. First, there are cases in which co-reference is manifest to the subject. In such cases, the thinker is warranted in trading on identity and thereby in inferentially exploiting the sameness of reference (c.f. Campbell 1987, 1994, 2002). Two thoughts that are related in this way are *de jure* co-referential. Second, there are cases in which two thoughts concern the same referent, but where the sameness of reference is not manifest to the subject. In such cases, the subject is not warranted in trading on identity. Such thoughts are *de facto* co-referential. This means that two pairs of thoughts that are referentially equivalent may nonetheless play different roles in cognition, depending on whether they are *de jure* or *de facto* co-referential. The question of how to account for the difference between *de jure* and *de facto* co-reference has been at the centre of many debates within philosophy of language and mind since Frege (1892). The account of samethinking I offer in this paper claims that the difference between *de jure* and *de facto* co-reference is to be understood in terms of a particular kind of irreducible relation – *the pointer relation* – that holds at the level of representational vehicles.<sup>2</sup>

We may distinguish between two classes of views about the difference between *de jure* and *de facto* co-reference. On the one hand, there are those who hold that the difference is to be accounted for in terms of intrinsic representational properties, i.e. properties that do not concern relations to other representations.<sup>3</sup> I will call such views *intrinsicist* views:

**Intrinsicism:** The difference between *de jure* co-reference and *de facto* co-reference is to be accounted for in terms of intrinsic representational properties.

---

the type/token distinction is, and (ii) the notions cannot be straightforwardly applied within the framework I will develop in this paper. Further, in this paper I focus on singular thoughts, i.e. thoughts that concern a single entity in an unmediated way. When I talk about reference, I thus have in mind the relation between a mental representation and the entity it represents rather than linguistic reference.

<sup>2</sup> By this, I mean that the relevant relations are *representationally* primitive. That is, the relations cannot be reduced to sameness or difference in intrinsic representational features of concepts and thoughts.

<sup>3</sup> Importantly, on this construal having a reference is considered an intrinsic representational feature of concepts on views that hold that having a reference does not depend on relations to other representations. Exceptions include views according to which reference is determined holistically, i.e. not independently of the content of other representations.

On the other hand, there are those who hold that samethinking cannot be accounted for purely in terms of intrinsic representational features (c.f. Fine 2007, Pinillos 2011). According to such views, the difference between *de jure* and *de facto* co-reference is essentially a matter of relational representational features. Whether or not two thoughts are *de jure* co-referential depends on how the thoughts are related, and such relations cannot be reduced to sameness or difference of intrinsic representational features. Such views are often labelled *relationist* views:

**Relationism:** The difference between *de jure* co-reference and *de facto* co-reference is to be accounted for in terms of irreducible relational representational features.

I will motivate a version of relationism by criticising a particular strand of intrinsicist views, according to which samethinking is to be accounted for in terms of co-location of information within mental files (c.f. Perry 1980, Recanati 2012, 2016). I argue that on the assumption that we want a unified account of samethinking, such views collapse into relationism. I then develop a new account of samethinking in terms of pointer relations.

## 2. De Jure Co-Reference: The Phenomenon

Two sets of thoughts may be referentially equivalent but still play different roles in cognition. Assuming Millian/Russellianism about mental content (as I will throughout this paper),<sup>4</sup> if two inferences are referentially equivalent then necessarily if one of them is truth-preserving so is the other. Even so, one of the inferences may turn out to be rationally warranted while the other is not. That is, the fact that a certain inference is truth-preserving does not guarantee that the inference is rationally warranted. Whether or not an inference is rationally warranted depends, at least partly, on the way the referent is represented in thought. The two following sets of beliefs are referentially equivalent, but only a

---

<sup>4</sup> More on Millian/Russellianism and my motivation for assuming this view in section 2 in the introduction chapter.



subject holding the first set of beliefs can combine the beliefs in a generalization: [CICERO WAS ROMAN, CICERO WAS AN ORATOR], [CICERO WAS ROMAN, TULLY WAS AN ORATOR]. For someone who only has the second pair of beliefs, the generalization that someone is a Roman orator would only be rationally warranted if she has the further belief CICERO = TULLY. In contrast, a generalization from the first set of beliefs does not require any identity judgements, since the identity is manifest to the thinker: In the case of *de jure* co-reference, the thinker knows a priori that the relevant terms co-refer.<sup>5</sup>

To see the importance of the phenomenon at hand, imagine if individuals were never rationally warranted in trading on identity of co-reference. What if all rational inferences essentially required identity judgements? In the case considered above, a thinker who holds the first set of beliefs would have to make a judgement of the form CICERO = CICERO in order to rationally make the generalization. But if this was the case we encounter problems when trying to explain how this can have any effect on the inference: How can the subject inferentially exploit this information if she can't trade on identity of the concepts occurring in the identity statement and the concept occurrences in the initial beliefs? If there were no trading on identity we would have to account for this in terms of yet further identity judgements, but this would just generate further problems of the same nature. Hence, accounting for all cases of knowledge of co-reference in terms of identity judgements is hopeless. There must be cases in which we can simply trade directly on identity of co-reference and these are cases where the co-reference is manifest co-reference, i.e. *de jure* co-reference.<sup>6</sup>

In the case of natural language the paradigmatic case of *de jure* co-reference is the case of anaphora. Someone who fully understands the sentence "Cicero admired himself" cannot fail to know that "Cicero" and "himself" refer to the same individual. In contrast, someone may fully understand the sentence "Cicero admired Tully" without knowing that "Cicero" and "Tully" refer to the same individual. However, as Lawlor (2002) points out, it is doubtful whether there is

---

<sup>5</sup> In Chapter 2 I will qualify this claim. For now, I stick to Recanati's specific understanding of *de jure* co-reference. Nothing in this paper hangs on this.

<sup>6</sup> See Campbell (1994) for more on this.

anything that plays the role of anaphora in thought. Why think that we use anaphoric concepts rather than simply reuse the relevant non-anaphoric concept? The latter makes for a much more economical theory of mental representations. In his (2016), Recanati suggests that whenever there is anaphora in natural language, this is to be accounted for in terms of sameness at the conceptual level: "The expressions 'John' and 'he' are not the same, in the ordinary sense of 'expression', but they are associated with the same conceptual representation, and that is what coindexing indicates" (2016, 9).<sup>7</sup> Hence, on this view, the phenomenon of anaphora in language is not mirrored in thought. Instead it must be explained in terms of a recurrence of the same concept (i.e. file) in thought.

Even if we do not find anaphora in thought, there is still something worth labelling *de jure* co-reference: As we have seen, in order for rational reasoning to get off the ground, we must allow for cases in which sameness of reference is manifest to the thinker. At the same time we have seen that sameness of reference is not enough to guarantee that a thinker will know that two concepts are co-referential. Hence there must be something more to the nature of thought beyond reference that explains why, in certain cases, co-reference is manifest while in other cases it is not. In this paper I suggest a novel way of understanding what this 'something more' can plausibly be. But before laying out my own account, I will critically assess an alternative account that has received a lot of attention in recent years, namely the mental file theory.

### 3. De Jure Co-Reference and Mental Files

According to the mental file framework, *de jure* co-reference is explained in terms of information clustering within mental files: All pieces of information contained within a single file are *de jure* co-referential. What file a given piece of information is stored in wholly depends on the acquaintance relation (or

---

<sup>7</sup> This is put in contrast with Fiengo and May (1996, 1998), according to whom anaphora involve retokening a linguistic expression of the same type.

*epistemically rewarding* relations, as Recanati (2012) calls them) through which it was gained. All pieces of information gained through the same acquaintance relation go into the same file and thereby become *de jure* co-referential. Being gained through a given acquaintance relation counts as an intrinsic representational feature. This is because being gained on the basis of a given acquaintance relation does not depend on relations to other representations. Likewise, being stored in a given file does not depend on relations to the other pieces of information that may be in that file. Since the explanation of *de jure* co-reference is given in terms of sameness of intrinsic representational features, the mental file picture is an intrinsicalist view about *de jure* co-reference.

A mental file contains information in the form of predicates taken by the subject to be satisfied by the referent of a file. So, for instance, my mental file about Cicero contains information such as ‘was Roman’, ‘was an orator’, and so on. On this picture, two pieces of information, *i* and *j*, are *de jure* co-referential if “*i* and *j* occur in the same file without the benefit of a prior judgement of identity” (Recanati 2012, 95).<sup>8</sup> This means that the fact that the two pieces of information ‘was Roman’ and ‘was an orator’ are stored within one and the same file is what warrants an inference directly from this to the conclusion that one and the same person was a Roman orator. So, on this picture every piece of information that is taken to concern the same referent is stored within the same file.<sup>9</sup>

In contrast, two pieces of information being stored in distinct files is an indication that the information is taken to concern two different referents. Let’s say that instead of having both ‘was Roman’ and ‘was an orator’ stored in the

---

<sup>8</sup> To say that two pieces of information are *de jure* co-referential might seem like an odd way of expressing the point, but the idea is simply that in such cases the fact that the two pieces of information regard the same referent is manifest to the thinker. I borrow this terminology from Recanati (2012, 2016).

<sup>9</sup> This picture becomes a bit more complicated on Recanati’s account, according to which two pieces of information might be taken to concern the same referent as a result of an identity judgement and still they need not be stored in the same mental file. In such cases there is an operation on files that allows for information to be shared, but that does not need to entail that every piece of information is actually stored in the same file. For more on this operation, called *linking*, see Recanati (2012, 42-53). For an alternative view on how to account on informative identity judgements on the mental file view see Lockwood (1971), Strawson (1974) and Recanati (2016). According to such views, informative identity judgements are to be accounted for in terms of *merging* of two (or more) files. For criticism of the merge model, see Millikan (1997, 508).

same file, I have the first predicate stored in the CICERO file (i.e. a singular file referring to Cicero) and the second predicate stored in another file, namely my TULLY file (i.e. another singular file that also refers to Cicero). Let's also assume that I am not aware that Cicero is Tully. Now, since the two pieces of information are stored in distinct files, I am not warranted in inferring from these two pieces of information alone that someone was a Roman orator. The information being stored in distinct files entails that the two pieces of information are not accessible from one and the same file: "to say that there are two distinct mental files is to say that the information in one file is insulated from information in the other" (Recanati 2012, 42). This is why one cannot, without violating rational norms, trade on identity of two pieces of information that are stored in distinct files. This is also what accounts for the possibility of rational individuals ascribing contradictory predicates to the same individual: Since a difference in files makes it the case that the information is insulated, there is nothing that prevents me from storing the predicate 'was Roman' in the CICERO file and 'was not Roman' in the TULLY file, since the contradiction is not manifest to me. On this picture, then, "information integration and inferential exploitation of information only takes place within files" (Ibid., 43).

#### 4. De Jure Co-Reference of Two-Place Predicates

The cases we have looked at thus far have been cases involving only one-place predicates. However, some of the information we gain about the reference of a file will be relational, and hence some of the information within the files will be relational predicates. In the case of two-place predicates, for instance, the information predicates a relation between the referent of the file in which the information is stored and some other object or individual.<sup>10</sup> Consider the thought CICERO LOVED CAESAR: In this case, the thought predicates about Cicero that he loved Caesar. If someone comes to form this belief, the predicate 'loved Caesar' will go into that individual's CICERO file. Now, in many cases a subject will believe

---

<sup>10</sup> In particular, I'll focus on relational predicates that contain two (or more) singular concepts, as opposed to predicates that contain a singular concept and a general concept (such as 'Lisa likes cats').

multiple relational predicates concerning the same individuals. For instance, in addition to having the predicate ‘loved Caesar’ in her CICERO file, someone might also have the predicate ‘killed Caesar’ in her CICERO file, if she believes that Cicero killed Caesar. According to the mental file framework, the fact that the two occurrences of CICERO in the thoughts CICERO LOVED CAESAR and CICERO KILLED CAESAR are *de jure* co-referential is explained by the concept occurrences being associated with the same file. Further, the two pieces of information being taken to concern the same individual, Cicero, is explained in terms of the information being located within the same file: The co-location ensures the coordination of the first relata (i.e. CICERO and CICERO).

However – and this is where the problem arises – we also need to explain why the two occurrences of CAESAR are *de jure* co-referential, and why the two pieces of information are taken to concern the same individual, Caesar. It is clear that they are *de jure* co-referential, because the subject need not make the explicit judgement that CAESAR (figuring in the first piece of information) and CAESAR (figuring in the second piece of information) are co-referential in order to make the inference that there is an individual such that he was both loved by Cicero and killed by Cicero – or simply that there is an individual such that he was both loved and killed by one and the same person. How are we to explain this phenomenon on the mental file framework? We clearly cannot account for this in terms of the occurrences of CAESAR being located in the CICERO file; appeal to the CICERO file can only explain the CICERO occurrences being *de jure* co-referential. On the current picture, co-location within mental files can only explain *de jure* co-reference of the first relata (e.g. the two occurrences of CICERO), but not of the second relata (e.g. the two occurrences of CAESAR) of two-place predicates. The CICERO occurrences are related to the CICERO file in this specific way in virtue of being located in that file, whereas we do not have a similar relation that we can appeal to when it comes to the occurrences of CAESAR.

In his (1980) Perry foreshadows this worry for the mental file picture. In offering an account of singular thought in terms of an analogy with files, he says that “in the analogy I have presented, there is really no provision for handling [relational] predicates” (Perry 1980, 20). However, without spelling out any

details about the nature of this problem he continues by saying that he would like “to end by pointing out how considerable relational information can nevertheless be handled in such a system, for this is a point that seems importantly related to the study of what it is to think from a position in the world” (Ibid.). The relations Perry has in mind are the ones that hold between the observer and the object that is observed. For instance, if one of your files contains the predicate ‘x sits to the left’, what you believe is that x sits to *your* left. Having this sort of information in a file thus involves predicating a relation between the referent of the file and yourself. Perry’s response is to say that the fact that the file is *your* file explains any relation the referent may bear to you. His explanation thus accounts for a specific class of relational predicates; namely the ones having oneself as one of the relata.

However, while this may explain this specific class of relational predicates, Perry’s explanation is not applicable to the case at hand. There is nothing in the nature of either CICERO or CAESAR that allows us to reduce the relational predicates to one-place predicates in the same way we may when the thinker herself figure as one of the relata in the particular way addressed by Perry. In what follows I will argue that in order to give a full account of relational predicates the mental file theorist is forced to postulate a relation between pieces of information across distinct files. In our case, she must say that there is a certain relation between the occurrences of CAESAR stored in the CICERO file, and the CAESAR file itself that accounts for the concept occurrences being *de jure* co-referential. In the next section I aim to show that this relation cannot be explained in terms of information clustering within mental files. If I am right, we need something over and above the notion of information clustering in order to give a full account of co-reference *de jure*.

### 5. Information Distribution in Mental Files

Since co-location of information within one mental file is not sufficient to explain *de jure* co-reference in cases of relation predicates we need to appeal to

something more in order to give a full account of *de jure* co-reference. The question, then, is what this ‘something more’ can plausibly be on the mental file framework. Perhaps understanding the way in which information in the form of relational predicates is distributed in the files can be of help in this enquiry. After all, understanding the information distribution of non-relational predicates helps in accounting for *de jure* co-reference in such cases: The information being clustered together in the same file explains why the information is taken to concern the same individual without a prior judgement of identity. In his (2012), Recanati says that in the case of relational predicates information “is *shared* between two files” (Recanati 2012, 50). Exactly what this *sharing* of information amounts to is unclear, but he mentions two possibilities: One possibility is that information in the form of relational predicates is duplicated into all the relevant files, thereby making the information accessible from each of them. The other possibility is to say that the information is not duplicated, but instead we introduce a *pointer* into one of the files that takes us to the file in which the information is stored. I return to the latter alternative and the notion of a pointer in section 6, but first I will focus on the first option and argue that simply duplicating the information is not sufficient to explain *de jure* co-reference of relational predicates.

According to the duplication strategy, the information in question is stored in both the CICERO file and the CAESAR file. So the CICERO file will contain the predicates ‘loved Caesar’, ‘killed Caesar’ while the CAESAR file will contain the predicate ‘is loved by Cicero’, ‘was killed by Cicero’ and so on:

<b>Cicero</b>	
... loved Caesar	
... killed Caesar	

<b>Caesar</b>	
... was loved by Cicero	
...was killed by Cicero	

*Fig. 1*

The subject may then inferentially exploit the information stored in each file. From the information in the CICERO file she can infer that one and the same person both loved Caesar and killed Caesar. From the information in the CAESAR file she can infer that one and the same person was loved by Cicero and killed by Cicero. This picture looks a lot like the original picture involving monadic predicates; the only difference is that in the case of relational predicates the information in question goes into multiple files. The duplication strategy, then, does not really add much to the original picture. In what follows I will argue that as a result, the strategy on its own does not provide us with a framework capable of handling *de jure* co-reference of relational predicates.

Consider again the inference from ‘Cicero loved Caesar’ and ‘Cicero killed Caesar’ to ‘someone was loved and killed by one and the same person’. In the previous section we saw that from the information stored in the CICERO file the thinker can infer that (i) someone both loved Caesar and killed Caesar. Likewise, from the information in the CAESAR file she could infer that (ii) someone was both loved by Cicero and killed by Cicero. However, on the current picture she cannot infer from these two pieces of information alone that (iii) someone was both loved and killed by the same person. Remember that on the mental file framework what warrants trading on identity is co-location of information in files. Even though the two occurrences of CAESAR are located in the same file they are not related to the same file *in the right way*. It cannot be that co-location of information guarantees co-reference of the concept occurrences that are not themselves associated with the file in which the information is stored. To see this, consider having a further piece of information in the CICERO file, namely ‘Cicero feared Cleopatra’. If co-location of information guaranteed co-reference amongst concept occurrences that are not associated with the file in question, the two occurrences of CAESAR as well as the occurrence of CLEOPATRA would be *de jure* co-referential, and the subject would be warranted in inferring that someone loved, killed and feared one and the same person. But this is clearly wrong. Hence someone cannot on the current picture infer (iii) without judging that Cicero (figuring in the first piece of information) is Cicero (figuring in the second piece of information) or that Caesar (figuring in the first piece of



information) is Caesar (figuring in the second piece of information). But we have already established that a thinker can infer directly from the information ‘Cicero loved Caesar’ and ‘Cicero killed Caesar’ that someone was loved and killed by the same person, without having to make any kind of identity judgement. On the current picture, the mental file framework fails to account for this.

The problem with the duplication strategy is that it does not postulate any relation between the two occurrences of CAESAR in the CICERO file, nor does it postulate any relation between the occurrences of CICERO in the CAESAR file. Further, there is nothing in this story that explains how the information in the CICERO file relates to the information in the CAESAR file or that file itself (and vice versa). Recall that on the mental file framework information integration and inferential exploitation of information only takes place within files (unless you have two files that are linked as a result of an identity judgement): “Exploitation of information is blocked if the relevant information is distributed in distinct files, for then, there is no presupposition that all the information derives from the same object” (Recanati 2012, 43). Since on the current picture, the information is stored in distinct files, and there is no further mechanism that relates the pieces of information in one of the files to pieces of information in the other (see *Fig.1*), we simply do not have the resources needed to give a satisfactory account of *de jure* co-reference of relational predicates. Duplication of information can only explain why two pieces of information within one and the same file comes to be *de jure* co-referential, and hence the duplication does not add anything of explanatory interest to the initial picture of *de jure* co-reference. Since, as we have seen (c.f. section 4), the initial picture is insufficient for explaining *de jure* co-reference of relational predicates, the duplication strategy fails to give a full account of the phenomenon.

In the next section I turn to the alternative story about information distribution of relational predicates, namely the story that appeals to the notion of a pointer. I will argue that this strategy is on the right track, but that it ultimately leads to the conclusion that mental files are redundant when it comes to explaining *de jure* co-reference. Since this is arguably the central purpose of the file framework, this is a significant blow to the theory.

## 6. Introducing Pointers

The problem with the duplication strategy was that on this picture there is nothing that relates the occurrences of CAESAR in the CICERO file to each other, or to the CAESAR file itself. Once this observation is made, appealing to the second way of representing information sharing suggested by Recanati seems more promising. According to this suggestion, information sharing amounts to “storing the information in a single file and introducing into the other file a pointer to the first file so as to make the shared information accessible from the second file” (Recanati 2012, 50). Again, this is an explanation of how the information is distributed in the files. However, introducing the notion of a pointer that indicates a relation between files might be exactly what we need in order to explain the cases we have looked at.<sup>11</sup> Returning to our example, according to this view, the information ‘Cicero loved Caesar’ and ‘Cicero killed Caesar’ would only be located in the CICERO file and not in the CAESAR file. But even though the information is not stored in the CAESAR file, it is still accessible from that file in virtue of the pointer that takes us from the CAESAR file to the CICERO file. The current picture, then, looks like this:

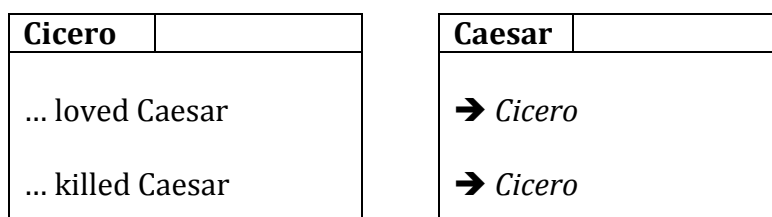


Fig. 2

<sup>11</sup>Recanati doesn't discuss the nature of the pointers, but in a footnote he refers to Hendriks' (2002), in which Hendriks discusses Vallduví's (1992) theory of information packaging. According to him, the pointer-mechanism is “'much more efficient' than a straightforward multiple recording of information on cards” (Hendriks 2002, 80). Note that Vallduví calls the pointer 'a linking mechanism', but this is not what Recanati has in mind when he uses the term 'linking'.

This picture avoids some of the problems posed for the duplication strategy. We now have a direct relation between the two files, and this helps us overcome the problem of insulation of information exploitation in files.

However, while this is an improvement, the picture is still not quite what we need. One problem with the current notion of a pointer is that the pointer seems to be a mechanism that connects files rather than pieces of information. In our case, the pointers signal that there is *some* information about the referent of the CAESAR file located *somewhere* in the CICERO file. But then the question is this: Given that our CICERO file may contain various further predicates, some of which may not concern Caesar, how do we know which predicates are the ones related to the CAESAR file?

One possibility is that the pointer tells us to go to a general file and then the cognitive system searches through all the information in that file until it finds some predicate(s) concerning the referent of the file in which the pointer is located. However, there are two problems with this strategy: First, this story seems too inefficient; it seems highly unlikely that in order to come to the conclusion (iii) *someone was killed and loved by the same person* (without making any identity judgements of the form CAESAR (figuring in the first piece of information) = CAESAR (figuring in the second piece of information)), the cognitive system would need to run through all the information stored in the CICERO file. Second, even if we allow for such inefficiency, we would have no way of explaining how the system would be able to recognize which pieces of information concern the referent of the file in which the pointer is located. Keep in mind that on the mental file framework sameness of mental representations is not accounted for in terms of sameness in Mentalese typography or the like, but rather in terms of sameness of mental files.<sup>12</sup> On the current picture we have no story to tell as to how the cognitive system is able to recognize that the two occurrences of CAESAR in the CICERO file stand in the relevant relation to the CAESAR file itself.

---

<sup>12</sup> Alternatively one might say that on this picture sameness in Mentalese typography is explained in terms of sameness of files.

All of this indicates that in order for the notion of a pointer to be of any help, the pointer must be such that it takes us directly to the relevant information, rather than to the general file in which the information is stored. Luckily there seems to be no principled reason why we cannot adopt this notion of a pointer on the mental file picture. The current picture, then, is this:

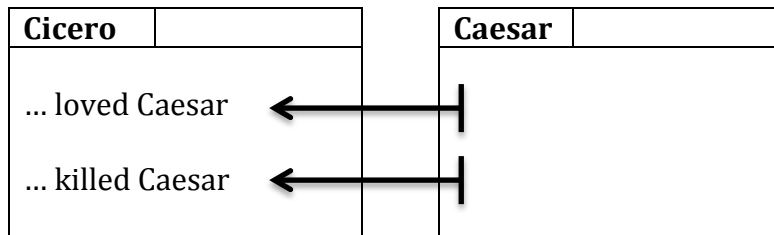


Fig. 3

The pointers indicate what information in the other file involves the referent of the file in which the pointer is based. Since there can be various different pieces of information stored in another file that concerns the referent of the file in which the pointer is based, we need one pointer for each piece of information.

While this is a further step in the right direction there is one further problem with the current notion of a pointer. According to the notion at work, a pointer is something that takes us from the file associated with the second relatum of two-place predicates to the information stored in the file associated with the first relatum. In the case at hand (Fig. 3), if we go to the CAESAR file we find something along the lines of a command that tells us to go to the CICERO file. In contrast we find no command similar to this in the CICERO file. That is, there is nothing in the CICERO file that tells us what file the two occurrences of CAESAR are associated with. Moreover, there is nothing in the CICERO file that tells us that the two occurrences of CAESAR are related to the *same* mental file, and thus that they are co-referential *de jure*. Just the fact that the two occurrences of CAESAR are being pointed to is not sufficient for them being *de jure* co-referential. After all, the thinker might have further relational predicates in her CICERO file. For instance, if she thinks that Cicero admired Aristotle, she would also have the

information 'admired Aristotle' in her CICERO file. In this case, there would be a pointer anchored in the thinker's ARISTOTLE file pointing to the relevant piece of information in the CICERO file. Hence, the occurrence of ARISTOTLE would also be pointed to, but the occurrence of ARISTOTLE is not *de jure* co-referential with the two occurrences of CAESAR that are also being pointed to.

The problem is that there is nothing in the CICERO file that tells the cognitive system which of the pointers are anchored in the same or different files. The one-way pointer can only do its job after we have established a relation between the CAESAR occurrences in the CICERO file and the CAESAR file. But in order to establish this relation we need something that takes us from the file in which the information in question is stored (i.e. the CICERO file) to the file associated with the second relatum (i.e. the CAESAR file). Only if we trace the pointer back to the file in which it is anchored (i.e. the CAESAR file) does it become clear that the two occurrences of CAESAR are related to each other, in that they stand in a pointer relation to the same mental file. In order to give a full account of *jure* co-reference, then, we need a notion of a pointer according to which the pointer goes both ways; the pointer must not only take us from the file associated with the second relatum to the file associated with the first relatum, but it must also take us from the file associated with the first relatum to the file associated with the second relatum:

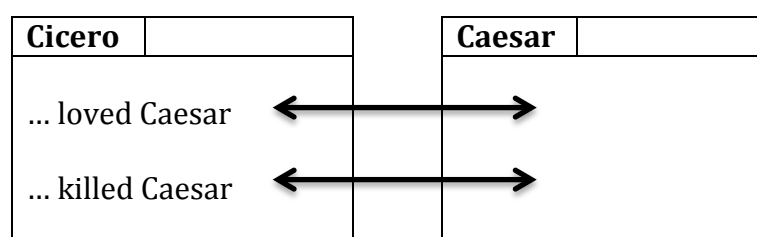


Fig. 4

The resulting picture, then, is one in which *de jure* co-reference is not simply explained in terms of information clustering in files, but also in terms of relations between information in distinct files. Even though information integration and

inferential exploitation of information only take place within files, it essentially depends on relational mechanisms across files. In the next section, I will argue that as long as we have the notion of a pointer relation, we do not need the notion of a file to account for *de jure* co-reference.

### 7. De Jure Co-Reference and Pointer Relations

We have seen that we cannot give a full account of *de jure* co-reference simply in terms of information clustering within mental files. This is because, in the case of relational predicates, the information is clustered in two distinct files in such a way that we cannot inferentially exploit the information without introducing further machinery. In order to account for *de jure* co-reference of relational predicates, then, we need a further notion; such as the notion of a pointer. A pointer, I suggest, is to be understood as a relation that holds between mental representations, and whenever this relation holds the relata are *de jure* co-referential. The mental file theorist, as I have argued, needs the notion of a pointer in addition to the notion of a file. She would have to say that while *de jure* co-reference of one-place predicates is explained in terms of co-location within files, *de jure* co-reference of two-place predicates is to be explained in terms of the specific relations that hold *across* files. But, then, there seems to be a pressing question: If we are to explain *de jure* co-reference in terms of pointer relations in the case of relational predicates, why think that we should explain *de jure* co-reference of one-place predicates in terms of co-location of information within mental files? If we can explain *de jure* co-reference of one-place predicates in the same way we explain *de jure* co-reference in the case of relational predicates, we do not need the mental files to explain *de jure* co-reference in the simpler cases.<sup>13</sup>

---

<sup>13</sup> One might think that even though *de jure* co-reference is not to be explained in terms of mental files, there is still some job left for the mental files to do. I think this is not the case; whatever the notion of mental files can do, we can do in terms of pointers. It is beyond the scope of this paper to give an argument to this effect. For the purposes of this paper, we may allow that an appeal to the notion of mental files is warranted in some cases. The important point is that when it comes to *de jure* co-reference we must abandon an explanation in terms of files. If anything, the mental file theorist has it the wrong way around: the *de jure* co-reference relation is what explains certain aspects of mental files, c.f. Fine: “mental files should be seen as a device for keeping track

The mental file theorist might respond by saying that we do, at the very least, need files to ‘anchor’ the pointers. That is, she might say that we cannot have pointers all the way down – the pointers must be anchored in something more fundamental. I agree that pointers must have anchoring points, but I do not agree that such anchoring points must, or should, be characterized as mental files. As long as we have pointers to account for the *de jure* co-reference relation, files are no longer doing any real work in the explanation of this phenomenon. A mental file is a metaphysically demanding notion: In postulating mental files we postulate small containers for each and every one of our mental representations (i.e. concepts), and each of the containers contain pieces of information – some of which involve occurrences of terms associated with different files.

I suggest that we give up on the idea of mental containers and instead we appeal to a notion of a mental *tag*, or the like. A tag, I suggest, is to be understood as a meeting point for pointers. Mental tags have two roles relevant to the explanation of *de jure* co-reference: (i) mental tags are meeting points for pointers, and (ii) mental tags have semantic contents. They are mental representations, building blocks of thoughts. As such they should be understood as representational vehicles. Whether or not two tags are *de jure* co-referential wholly depends on whether or not they are connected by a pointer. *De facto* co-reference occurs whenever two thoughts concern the same referent, but where the relevant tags are not connected by pointers.

I suggest, then, that the reason why we can infer from CICERO WAS ROMAN and CICERO WAS AN ORATOR that someone was a Roman orator without an additional premise of the form CICERO = CICERO is that there is a pointer between the two occurrences of CICERO relating them in such a way as to explain them being *de jure* co-referential. In our case of information in the form of two-place predicates, the first and second relata are not connected by pointers. We may illustrate the pointer picture thus (although I will qualify this shortly):

---

of when objects are coordinated (represented as the same) [i.e. *de jure* co-referential] and, rather than understand coordination [i.e. *de jure* co-reference] in terms of mental files, we should understand the workings of mental files in terms of coordination [i.e. *de jure* co-reference]” (2007, 68).

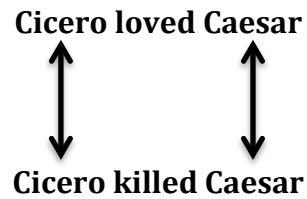


Fig. 5

The picture I'm suggesting is metaphysically sparse. We only need the notion of a tag and the relations between these, in addition to a simple Millian/Russellian picture of the propositional content of thoughts. At the same time, the framework is explanatorily powerful in that it accounts for the phenomenon of samethinking just as well as the mental file picture arrived at in the end of section 6.

The pointer relation is representationally primitive. That is, the relation that accounts for *de jure* co-reference cannot be reduced to sameness of type of tags, or the like. In particular, the pointer relation does not hold in virtue of sameness of intrinsic syntactic properties, such as sameness of Mentalese symbols. On the pointer picture, the intrinsic syntactic properties of mental representations play no role in determining when two representations are – or fail to be – *de jure* co-referential. The framework has a minimum of metaphysical commitments, and this makes the framework much more flexible than theories that account for samethinking in terms of intrinsic representational features, including the mental file account. To see why this is a great virtue of the theory consider the following cases:

- 1) We were debating whether to investigate both Hesperus<sub>1</sub> and Phosphorus<sub>2</sub>; but when we got evidence of their true identity, we immediately sent probes there<sub>1,2</sub>.
- 2) Hesperus<sub>1</sub> is Phosphorus<sub>2</sub> after all, so Hesperus-slash-Phosphorus<sub>1,2</sub> must be a very rich planet.



In both cases we have two concepts that are *de jure* co-referential with the latter concept (i.e. ‘there’ and ‘Hesperus-slash-Phosphorus’), but that are not *de jure* co-referential with each other. In his (2011) Pinillos uses these cases to argue that *de jure* co-reference is a non-transitive relation. If this is correct, this creates huge problems for any theory that attempts to account for *de jure* co-reference in terms of identity relations of any kind (be it sameness of Mentalese type or mental files or the like). If *de jure* co-reference were a matter of identity of this sort, the relation would be transitive. Hence, if Pinillos is right, such intrinsicist views fails to give a general account of *de jure* co-reference.

Note that (1) is a case of anaphora. As mentioned earlier (section 2) it is controversial whether or not we have anaphora in thought – why not just use our non-anaphoric concepts twice? Those who hold this view might find (2) more convincing. In this case we do not have anaphora, but instead we have a slash-concept – the kind of concept that (in normal circumstances) results from an informative identity judgement. The idea is that after having made the judgement we get a new concept that is a merging of the two (or more) original concepts.<sup>14</sup> One may of course question whether there is such a thing as a slash-concept. But then one would have to give an account of why Pinillos’ argument has a strong intuitive appeal. The question of whether or not there are such things as slash-concepts requires further investigation. What’s important for this paper is simply that the pointer picture – unlike the mental file view – is compatible with either outcome: If it turns out that *de jure* co-reference is in fact a non-transitive relation, this would not pose any problems for the pointer picture, since on this framework we do not explain *de jure* co-reference in terms of identity relations.

Since mental tags are to be understood as meeting points for pointers, rather than Mentalese symbols or the like, a more correct way to illustrate the pointer picture is this:

---

<sup>14</sup> Such a view can be found in Recanati 2016.

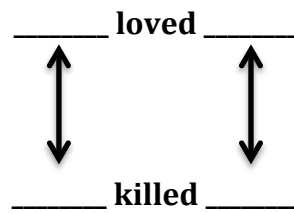


Fig. 6

The empty spaces represent (instantiated) mental tags. What matters for the explanation of *de jure* co-reference is merely whether or not the tags are connected by pointers. From the case illustrated, we can conclude directly that someone was both loved and killed by the same person as a result of the tags being related by pointers in this specific way. We may, then, illustrate the case of non-transitivity of *de jure* co-reference discussed above thus:

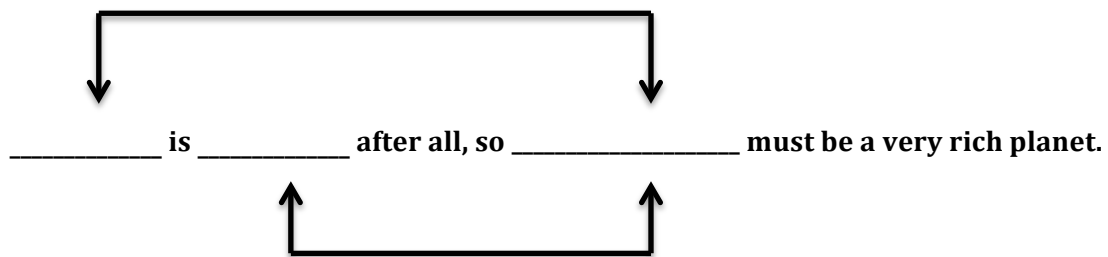


Fig. 7

The first and second mental tag stand in pointer relations to the last one (the slash-concept), but they do not stand in pointer relations to each other. On this picture, *de jure* co-reference is fully explained in terms of how the mental representations are related to each other.

A further virtue of the pointer picture is that it is capable of explaining *de jure* co-reference between all the different propositional attitudes and not just doxastic states. A mental file, recall, is by definition a cluster of information taken by the subject to be true about the referent. What this in effect means is that the mental file framework cannot give us an explanation of how a desire and a belief can

stand in the *de jure* co-reference relation. However, a subject's ability to combine beliefs and desires in a way that does not involve identity judgements is crucial for explaining behavioural dispositions. The problem does not arise for the pointer picture because it does not make a principled distinction between different attitudes. If you desire to read a book about Cicero and you believe that going to the library is a good way to read about Cicero these two thoughts involve *de jure* co-reference and combining them will result in you having a (rational) behavioural disposition to go to the library. The desire and the belief are *de jure* co-referential because they stand in a pointer relation.

The mental files theorist also has problems in accounting for reasoning involving doxastic attitudes other than belief. Consider suppositional reasoning. In such reasoning, the premises are not taken to be true by the subject. Yet they may enter into *de jure* co-reference relations – either to each other, or to information stored within mental files. Consider also deliberation about e.g. whether or not you want to read more about Cicero. In this case, it might be relevant to take into account your beliefs about Cicero. If so, it is crucial that the cognitive system is able to detect which beliefs are relevant for your deliberation, and this as well involves *de jure* co-reference.<sup>15</sup>

The pointer picture has no problem here. Again, it predicts that *de jure* co-reference may obtain across all kinds of propositional attitudes. It is a mistake to make a distinction between the different attitudes when it comes to *de jure* co-reference.

In the next section, I will show how the pointer picture accounts for cognitive significance. I will also compare the pointer picture to other theories according to which *de jure* co-reference is explained in terms of relational properties. In particular, I will point to differences between the pointer picture and the sort of *semantic* relationism found in Fine (2007).

---

<sup>15</sup> Similar issues also arise for the mental file account when it comes to accounting for *de jure* co-reference in imagining (cf. Ninan 2015).

## 8. The Pointer Picture, Cognitive Significance and Relationism

In recent years there has been a trend in the debate about *de jure* co-reference to appeal to relational properties of concepts and words in order to account for the phenomenon (e.g. Fine 2007, Pinillos 2011). On such views, the relevant relational aspects are primitive, and thus cannot be accounted for simply in terms of intrinsic representational features. The pointer picture is to be seen in line with such relationist theories. On the framework I'm proposing there is nothing about the intrinsic nature of the two tags involved in thinking CICERO IS ROMAN and TULLY IS ROMAN that tells us whether or not the thoughts are the same. It is only once we see how the beliefs relate to each other (i.e. whether they are connected by pointers or not) that we can know whether or not it is in fact the same belief. In what follows, I will show how this framework explains how two referentially identical thoughts may nonetheless play distinct roles in cognition.

### *8.1. Cognitive Significance*

Pointers account for the role a given belief plays in thought. We can, therefore, account for the cognitive significance of different beliefs by appealing to pointer relations. Take for instance the two thoughts CICERO IS CICERO and CICERO IS TULLY. In the former thought, the two tags are connected by pointers, whereas in the second they are not. We may then explain the difference in informativeness in terms of how the beliefs interact with a person's other beliefs. That is, we may understand the *cognitive impact* of the two beliefs in terms of how they relate to one's already existing belief base. Cognitive impact is understood as the result of a new belief's being taken as input to an already existing base of beliefs (c.f. Fine 2007). Now, let's say that a given individual's belief base does not already include the belief CICERO IS TULLY. The two thoughts CICERO IS CICERO and CICERO IS TULLY will differ in how they relate to this individual's belief base, and thus in their cognitive impact. In the first case the two mental tags are themselves connected by pointers. When someone forms this belief, both tags

will enter into the exact same pointer relations: Both tags will stand in pointer relations to those beliefs in the belief base that the individual might express using the name 'Cicero'. As a result, the cognitive impact will be minimal: The individual will not be able to exploit the pointer relations in such a way as to gain new knowledge. In contrast, if the person formed the belief *CICERO IS TULLY*, the two mental tags would not be connected by pointers. The two mental tags would enter into different pointer relations. The first tag would stand in pointer relations to all beliefs in the belief base that the individual might express using the name 'Cicero', whereas the second tag would stand in pointer relations to all of the beliefs she might express using the name 'Tully'. In this case the individual may exploit the new pointer relations and draw new inferences. For instance, if she has the belief *CICERO IS ROMAN* and *TULLY IS AN ORATOR*, the pointer relations going between each of these beliefs and the new belief (i.e. *CICERO IS TULLY*) will warrant the conclusion that someone is a Roman orator.

Importantly, coming to know that Cicero is Tully does not have the cognitive result that occurrences of *CICERO* and *TULLY* enter into direct pointer relations. Pointer relations only hold in cases where the recognition of co-reference is not due to a prior identity judgement (cf. Recanati's claim that two pieces of information in a file are *de jure* co-referential only if their being located in the same file is not due to a prior judgement of identity (see section 2)). Instead, what happens in the case of informative identity judgements is that one may draw new inferences that one was previously not warranted in making. The identity judgement functions as an implicit premise and thereby rationally allows the thinker to draw new conclusions from premises that are only *de facto* co-referential. We see, then, that there is a relative difference in the cognitive impact of the two kinds of identity judgements. This, I suggest, is what the difference in cognitive significance of trivial and informative identity judgements amounts to.

Pointers also account for the possibility of rational individuals ascribing contradictory predicates to the same referent. Recall that on the mental file framework the explanation was that information stored in distinct files are insulated from each other. The explanation in terms of pointers appeals to

whether or not the beliefs stand in pointer relations to each other. In cases where a subject rationally believes two incompatible propositions, as in the case of someone believing that Cicero is bald and also that Tully is not bald, the two beliefs – or more specifically, the belief constituents – are not connected by pointers. If two beliefs are not connected by pointers, they are not taken by the subject to concern the same referent. If they happen to be co-referential this is not manifest to the thinker. This is why one may rationally believe of the same referent both that he is bald and that he is not bald. The explanation in terms of pointers does an equally good job of explaining such cases as does the mental file framework. But the pointer picture has the benefit of not having to appeal to identity relations (i.e. identity of files). As we have seen, theories that appeal to identity relations face problems in accounting for the possibility of *de jure* co-reference being non-transitive.

In the next sub-section I will point to similarities and differences between the pointer picture and other relationist accounts found in the literature, such as those of Fine (2007) and Pinillos (2011).

### *8.2. Different Versions of Relationism*

I take the main claim of relationism (about thoughts) to be that samethinking cannot be explained purely in terms of intrinsic features of concepts and thoughts, and that we need to take the relational aspects of thinking into account in order to give a full account of *de jure* co-reference. While relationists such as Fine and Pinillos hold that the relevant relations hold at the level of propositional content,<sup>16</sup> the picture I have suggested does not commit us to say that the relational properties relevant to the explanation of *de jure* co-reference in thought are part of, or affect, the propositional content of thoughts. It is compatible with the pointer picture to say that the two thoughts CICERO IS CICERO and CICERO IS TULLY have the same propositional content. What makes the cognitive difference is how the representational vehicles relate to a thinker's

---

<sup>16</sup> I set out Semantic Relationism in more detail in Chapter 2.

other representational vehicles. This is why the pointer picture can allow for a simple (i.e. non-relational) Millian/Russellian account of mental content. The pointer picture (or what one may call *Vehicle Relationism*) provides a relationist explanation of *de jure* co-reference but it does so without invoking a complex semantics. As far as I can see, when it comes to explaining *de jure* co-reference in thought, there is nothing the semantic relationist can do that can't be done on the pointer framework, and the suggested framework is compatible with a parsimonious semantics, which counts in its favour.

A notable feature of the pointer picture is that, if correct, contrary to what has often been assumed in the debate, the explanation of *de jure* co-reference is not fundamentally a matter of semantic properties. Rather, *de jure* co-reference is essentially a matter of relations between representational vehicles. Two thoughts such as HESPERUS IS HESPERUS and HESPERUS IS PHOSPHORUS may have the same content, but still play different roles in cognition. *De jure* co-reference is at the heart of many of the classical puzzles within the philosophy of mind, and the pointer picture predicts that trying to explain such puzzles in terms of semantic sameness or difference has it, at best, the wrong way around: The explanation is essentially non-semantic in nature.<sup>17</sup>

## 9. Conclusion

I have argued that the mental file theorist cannot account for *de jure* co-reference of relational predicates in terms of co-location in files. I considered the two strategies available to the mental file theorist: the duplication strategy and the strategy of appealing to pointer relations. I argued that the duplication strategy does not add anything of explanatory interest to the original picture. I then showed that the strategy of appealing to pointer relations undermines the mental file picture: I argued that the explanation of *de jure* co-reference in terms of pointer relations can be generalized to cases of monadic predicates. Finally, I argued that it is the notion of a pointer rather than the notion of a file that is of

---

<sup>17</sup> I develop the pointer picture further in Chapter 2.

explanatory interests in the case of *de jure* co-reference. As a result we should abandon talk of mental files in favour of a unified account of *de jure* co-reference in terms of pointer relations.

The positive view has several virtues. First, it is much simpler than the mental file framework: both views are in need of pointers, but on the view I've suggested we do not need any further apparatus to account for *de jure* co-reference. Second, the framework makes no principled distinction between the different propositional attitudes; whereas the mental file framework only accounts for *de jure* co-reference of doxastic states, the pointer picture can account for how the relation may obtain across all kinds of attitudes. Third, the suggested framework is highly flexible which allows for the possibility of *de jure* co-reference being a non-transitive relation. Finally, since pointer relations hold at the level of representational vehicles rather than at the level of content, appealing to pointers allows us to account for the Fregean data within a Millian/Russellian framework. The notion of a pointer is a minimal addition to the classical Millian/Russellian picture, but it is explanatorily powerful. I've suggested how it can account for the difference between trivial and informative identity judgements and for how someone may be rational in ascribing contradictory properties to the same referent. Both explanations essentially appeal to the way beliefs are related to each other and a thinker's already existing base of beliefs. The pointer picture provides a metaphysically sparse and unified account of samethinking by appealing to a notion that the mental file theorist is forced to posit, but it does so without invoking the notion of a mental file.

These considerations open up a new area of research, where the nature of the pointer relation becomes a central issue. On my framework the pointers are primitive, but there's more to say: When does a pointer come into existence? What sustains its existence? Further research on the pointer relation will yield new insights into the nature of samethinking.





## Vehicle Relationism: In Defence of Pointers

*A pair of thoughts may be such that whether or not they concern the same referent is directly manifest to the thinker. However, sameness or difference in reference is not always manifest to thinkers in this way. Manifestness of co-reference can be accounted for in terms of the notion of coordination relations: If co-reference is manifest, the thoughts are positively coordinated, while in cases where co-reference is not manifest, the thoughts are negatively coordinated. An important question, then, is this: How should we account for such coordination relations. In this paper I propose a novel account of coordination relations in thought. I argue that coordination is to be explained in terms of primitive relations having representational vehicles as relata. I call this view 'Vehicle Relationism'. I show that Vehicle Relationism provides solutions to some of the longstanding puzzles within the philosophy of mind. I also compare Vehicle Relationism to similar accounts and argue that the proposed account is superior.*

### 1. Introduction

It seems intuitively plausible that people can know the content of their own thoughts. In particular, the content of singular thoughts – thoughts that are about single individuals in a direct way – seems to be transparent to thinkers. For

instance, if you believe that Ringo Starr was the drummer for the Beatles, you have a particular person in mind and you believe of that particular person that he has a certain property.

It may then seem equally plausible that one can know whether or not two thoughts concern the same referent. After all, if you have a further belief that Ringo Starr is from Liverpool it will be obvious to you that the two thoughts concern the same referent, and you can use this information to infer that someone who was the drummer for the Beatles is from Liverpool. However, it is not the case that such sameness of reference is always transparent to thinkers.

Suppose you have a further belief that Richard Starkey was a member of the band Rory Storm and the Hurricanes. This thought concerns the same person that you believe to be the drummer for the Beatles and who was born in Liverpool. In this case, it is possible that you are not aware of this fact; you may fail to know that Richard Starkey and Ringo Starr are one and the same person. Even though you may know the content of the singular thoughts in isolation, you may fail to know whether or not the two thoughts concern the same individual. We see, then, that singular thought is only transparent to a certain degree: One may properly be said to know the referent of singular representations and still fail to know whether or not two singular representations concern the same referent.

Whenever two mental representations refer to the same thing, and this fact is manifest to the thinker without her having to make an explicit identity judgement, we say that the representations are *positively coordinated*.<sup>1</sup> The two tokens of RINGO STARR in the example considered above are positively coordinated since the thinker does not need to make an identity judgement in order to be warranted in inferentially exploiting the information. In contrast, the token of RICHARD STARKEY is *negatively coordinated* with the tokens of RINGO STARR, since the co-reference is not manifest to the thinker in the same way. In

---

<sup>1</sup> In this paper I assume that there are mental representations and that such mental representations are structured in such a way that they have representational proper parts and that these parts have reference. The content of a thought depends on the content of its constituent parts and how these are structured.

order for the thinker to be warranted in inferentially exploiting the co-reference of these mental representations she must make the identity judgement to the effect that Ringo is Richard. How should we account for such coordination relations?

Some have argued that there is more to the content of our thoughts than reference and structured propositions (e.g. Frege 1892). Such philosophers hold that the full content of singular representations is transparent to thinkers, and that only the two tokens of RINGO STARR have the same content in this respect. Although RINGO STARR and RICHARD STARKEY share the same referent, their overall content differ. In this way transparency of co-reference, and also the limits thereof, is explained in terms of the semantic content of our thoughts: The full content of thoughts is transparent, and this accounts for how we may know – or fail to know – that two representations are co-referential.

Traditionally, this line of thought is found within the Fregean tradition, and such philosophers hold that coordination relations are to be accounted for in terms of identity of intrinsic representational features of concepts and thoughts. Intrinsic representational features are those that can be stated without reference to other mental representations (c.f. Gray 2017). In recent years, however, some philosophers within the referentialist tradition have made a similar move: According to Fine (2007), there is more to the semantic content of thoughts than just classical Millian/Russellian propositions. However, he denies that this ‘something more’ is to be construed as a further layer of semantics a la Frege. Fine claims that the referentialist propositions themselves contain the relevant information. This information is of the form of primitive semantic relations that hold between the constituents of the proposition. So when you know a priori that Ringo is Ringo this is because you stand in a relation to a proposition in which the object occupying the first and second slot stand in such primitive semantic relations. In contrast, when you form the belief that Ringo is not Richard, you stand in relation to a proposition whose constituents do not stand in such semantic relations. On this view, then, coordination is accounted for in terms of primitive semantic relations.

In this paper I offer and develop a novel account of coordination relations.<sup>2</sup> I call this view *Vehicle Relationism*. After having spelt out the positive view in some detail I argue that we have reasons to prefer Vehicle Relationism to Semantic Relationism. In fact, I will argue that, even though Vehicle Relationism and Semantic Relationism are in theory compatible, when it comes to coordination in thought we have reasons to reject Semantic Relationism all together in favour of Vehicle Relationism.

Here is the plan for the paper. In section 2, I spell out the details of my positive view. In section 3, I show how Vehicle Relationism solves some of the central problems within the philosophy of mind pertaining to cognitive significance. In section 4, I compare the proposed view to some similar accounts found in the literature, such as Heck's (2012) *Formal Relationism* and Fodor's (1975, 2008) *Language of Thought Hypothesis*. In part 5, I compare Vehicle Relationism and Semantic Relationism and offer arguments to the effect that Vehicle Relationism provides a better framework for understanding coordination in thought. Finally, in part 6, I set out some objections to my positive account and respond to these worries.

## 2. Vehicle Relationism

We may distinguish between two classes of views about how to account for coordination relations: On the one hand, there are those that hold that coordination is to be accounted for in terms of intrinsic representational properties, i.e. properties that do not concern relations to other representations. I call such views *intrinsicist* views. Intrinsicism has been the dominant view of coordination throughout the 20<sup>th</sup> century.<sup>3</sup>

On the other hand, there are those that hold that coordination cannot be accounted for purely in terms of intrinsic representational features (c.f. Fine 2007, Heck 2012, Pinillos 2011). According to such views, coordination is

---

<sup>2</sup> This is a further development of the framework suggested in Chapter 1 of this thesis.

<sup>3</sup> For more on this, see the introduction chapter, section 3.1.

essentially a matter of relational representational features. Such views are called *relationist* views. The view developed in this paper is a version of relationism.

The basic idea of relationism is that coordination is to be understood in terms of primitive relational features (c.f. Gray 2017). That is, according to this view, coordination is essentially a relational phenomenon that cannot be reduced to sameness or difference in intrinsic representational features. A further question concerns the nature of these relations. In particular, what is the nature of the relata? According to Vehicle Relationism the relations in question have representational vehicles – i.e. non-semantic mental particulars – as relata. This is where Vehicle Relationism and Semantic Relationism part ways: According to the Semantic Relationists the relations are constituents of propositions. On this view, the relations have objectual constituents of propositions as relata. The core claim of Vehicle Relationism is this:

**Vehicle Relationism:** Coordination is to be accounted for in terms of primitive relations between representational vehicles.

In what follows I will elaborate on this view. There might be other ways of cashing out the core claim of Vehicle Relationism, but I do so in terms of *pointer relations* and *mental tags*.

### 2.1. *Pointer Relations and Mental Tags*

I use the notion of *pointer relations* to denote the relevant primitive relations. A pointer, I suggest, is to be understood as a relation that holds between representational vehicles. Whenever this relation holds the relata are positively coordinated.<sup>4</sup> Importantly, the pointer relations do not reduce to sameness of

---

<sup>4</sup>The notion of a pointer is originally found within the literature on information packaging in linguistics (c.f. Vallduví, 1992). This literature appeals to information clustering on *file cards* (c.f. Heim 1982, 1983) in order to model information storage. There the need for pointer relations arises in cases where information about the referent of a file card is stored within a different file. This may, for instance, take place in cases of relational predicates, where the information concerns two distinct individuals. In such cases the information is stored on one of the relevant file cards, while a pointer is introduced into the other, giving instructions to go to the other file.

intrinsic representational features of the relata. That is, it is not possible to determine whether or not two representational vehicles are positively coordinated merely by looking at the properties of the two vehicles in isolation. According to Vehicle Relationism two representational vehicles may share the exact same intrinsic properties – including their semantic content – and still fail to be positively coordinated.

I call the representational vehicles that are capable of entering into pointer relations *mental tags*. Together with pointer relations, mental tags are the building blocks of thoughts, and as such they can be understood as individual concepts. However, I propose that when it comes to the explanation of coordination there is nothing of interest to be learned from looking at the intrinsic features of the tags – only relational features are explanatorily interesting when it comes to coordination. Mental tags have two roles relevant to the explanation of coordination: (i) mental tags are meeting points for pointers, and (ii) mental tags have a semantic content. I say more about the semantic content of mental tags in the next section.

We now have the following characterization of coordination relations:

**Positive Coordination:** Two mental tags are positively coordinated if and only if they are connected by pointers.

**Negative Coordination:** Two mental tags are negatively coordinated if and only if they are not connected by pointers.

Whether or not two tags are positively coordinated fully depends on whether or not they are connected by pointers. Even if two tags refer to the same individual, they may fail to be connected by pointers. This is what explains the possibility of rational individuals having semantically contradictory beliefs. I return to this in section 3 where I use the pointer picture to solve some of the classical philosophical problems pertaining to the cognitive role of concepts.

---

In Chapter 1 I argued that we ought to give up on the file metaphor when it comes to thoughts, but I find the pointer relation helpful. Thus, since my framework does not appeal to file cards or the like, the notion of a pointer is understood slightly differently from how it is understood in the literature on information packaging.

## 2.2. Mental Content

Vehicle Relationism is primarily a view about representational vehicles. In theory the framework could be combined with a number of different accounts of mental content. I am, however, independently motivated to adopt a classical Millian/Russellian account of propositions. According to this view, propositions are understood as structured complexes having objects and relations as constituents.<sup>5</sup> Millian/Russellianism is the most parsimonious account of propositions. The main challenge for the account is to explain how it can be that two pairs of co-referential concepts may nonetheless differ in coordination relations. In section 3, I show how one can give a Millian/Russellian reply to this challenge if one combines this account of mental content with Vehicle Relationism. In the rest of this paper I will simply assume Millian/Russellianism about mental content.

On the picture I have in mind, the reference of a singular representation (i.e. a mental tag) is determined by the relation the representation bears to the referent (and not e.g. by the referent satisfying a set of descriptions). Whenever someone forms a belief as a result of standing in a direct or indirect relation to an object, the relevant mental tag refers to the object to which they are related.<sup>6</sup>

In the rest of this section I will draw some explanatorily interesting distinctions that follow naturally from Vehicle Relationism. In particular, I will say more about what the difference between manifest and opaque co-reference amounts to on this framework. In short, the difference turns out to be wholly a matter of whether or not the relevant mental tags stand in pointer relations. I also draw a distinction between two sorts of positive coordination, namely manifest and

---

<sup>5</sup> More on Millian/Russellianism in the introduction chapter, especially section 2.2.

<sup>6</sup> Here I have in mind something along the line of Recanati's (2012) notion of an *Epistemically Rewarding* (ER) relation. Such ER relations are to be understood as acquaintance relations broadly construed, so that a thinker may stand in ER relations to objects even in cases where the information is gained through i.e. testimony. Note that the Vehicle Relationist is not committed to this particular view of reference-fixing. The viability of the framework does not depend on the viability of ER relations or the like.



apparent co-reference. In this case, the difference is fully a result of a difference in content.

### 2.2.1. *Manifest Co-Reference and Opaque Co-Reference*

On the view I'm suggesting, propositional attitudes are *triadic* relations holding between individuals, representational vehicles and propositions. The idea is that individuals do not stand in direct relations to the propositional content of thoughts, but rather they stand in relations to such contents only indirectly, via representational vehicles.<sup>7</sup> This claim, while not uncontroversial, is widely accepted (e.g. Fodor 1975, 2008, Fodor & Pylyshyn 1988, Laurence & Margolis 1999, 2007).<sup>8</sup> It is possible, then, that a thinker may unknowingly believe the same proposition twice if she has numerically distinct representational vehicles that express the same proposition, *and these representational vehicles (i.e. mental tags) are not connected by pointers*. For instance, in the case of someone having the beliefs RINGO STARR IS FROM LIVERPOOL and RICHARD STARKEY IS FROM LIVERPOOL the thinker has two numerically distinct beliefs that both express the same proposition. The beliefs are distinct as a result of the fact that the relevant singular mental tags are not connected by pointers.

From the definition of negative coordination above (i.e. section 2.1.) we see that two beliefs may be negatively coordinated even in cases where they happen to be co-referential. This is how we account for the distinction between *manifest* and *opaque* co-reference on the proposed framework. The set of propositional attitudes that are manifestly co-referential is a subclass of the attitudes that are positively coordinated. The other subclass of positively coordinated propositional attitudes consists of attitudes that contain mental tags that are connected by pointers but that are not actually co-referential (I return to this in section 2.2.2.). The set of propositional attitudes that are opaquely co-referential

---

<sup>7</sup> I borrow the notion of a 'triadic relation' from Aydede's (2010) characterization of Fodor's LOTH.

<sup>8</sup> I motivate the need for representational vehicles when it comes to the explanation of coordination in thought further in section 5.2 below.

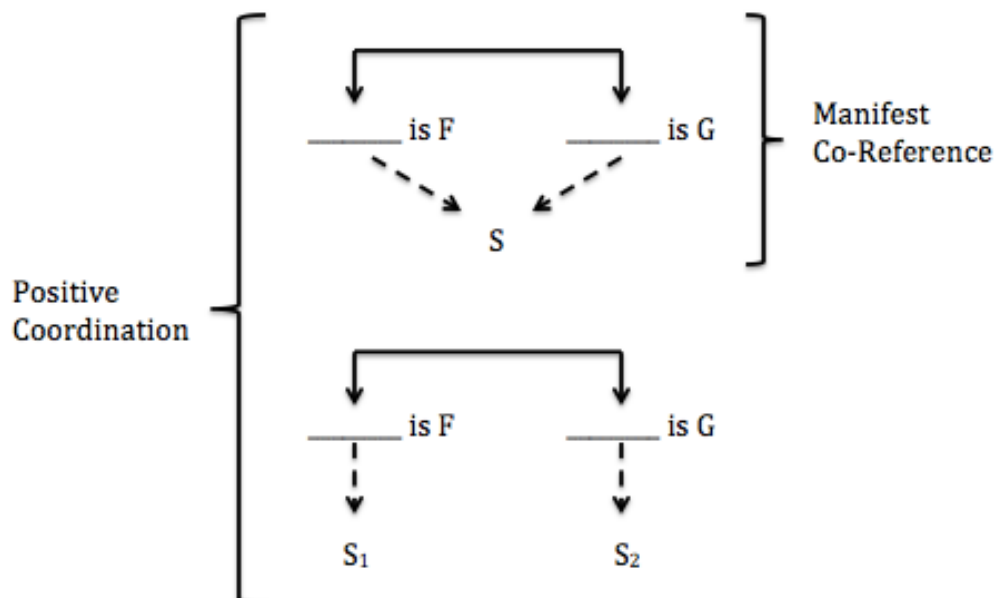
is a subclass of the attitudes that are negatively coordinated. The other subclass of negative coordination is the set of negatively coordinated beliefs that are not co-referential. Importantly, from the point of view of the thinker, there is no difference between cases of negative coordination in which the relevant mental tags refer to distinct entities and cases in which they refer to the same entity.

We may thus define the difference between manifest and opaque co-reference in thought as follows:

**Manifest Co-reference:** Two mental representations are manifestly co-referential if and only if they are positively coordinated and share reference.

**Opaque Co-reference:** Two mental representations are opaquely co-referential if and only if they are negatively coordinated and share reference.

We may illustrate these distinctions thus:



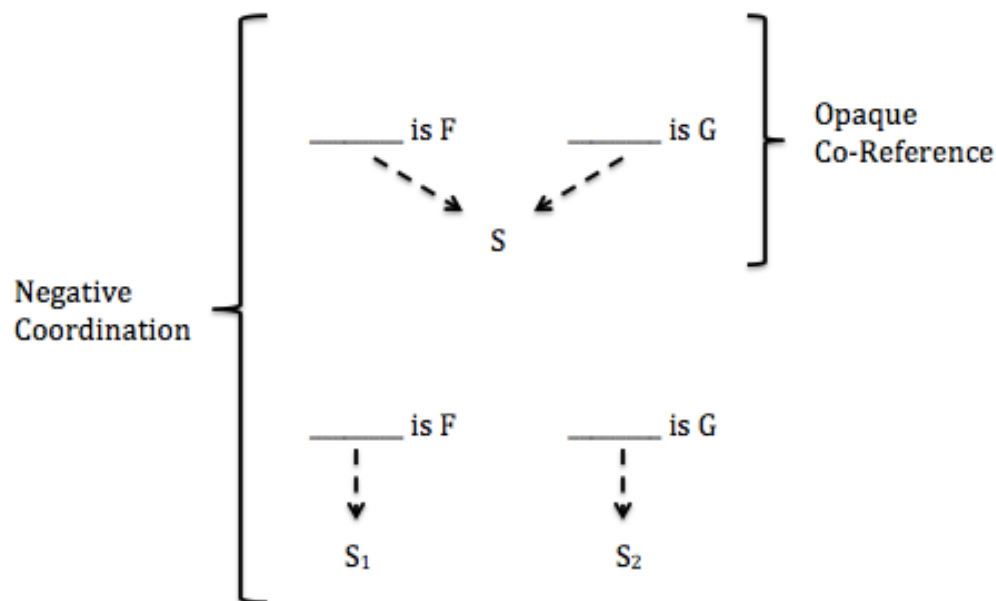


Fig. 1

The dotted arrows represent reference relations, while the unbroken arrows illustrate pointer relations. We see that on the proposed picture, positive coordination is a more general notion than manifest co-reference. In the bottom case of positive coordination there is only *apparent* co-reference, since the mental tags do not actually co-refer. I discuss the notion of apparent co-reference further in the next section.

### 2.2.2. Apparent Co-Reference

In the previous section we saw that there is another class of positively coordinated propositional attitudes in addition to attitudes that are manifestly co-referential. Such attitudes involve *apparent* co-reference. Cases of apparent co-reference include cases that are often called *inverse* Paderewski cases (c.f. Recanati 2012). In inverse Paderewski cases there is a subject who mistakenly takes two distinct referents to be the same. In such cases there are two tags that are connected by pointers but refer to distinct individuals.

Consider for instance Pam, who every morning watches the mail carrier through her window. She has various beliefs about the mail carrier, such as SHE IS ALWAYS ON TIME, SHE HAS BROWN HAIR, SHE WEARS COMFY SHOES and so on. Unbeknownst to Pam, however, there is not one but two mail carriers who happen to be twins. The twins work every other day and Pam never sees them at the same time. Now Pam's beliefs are connected by pointers and so they are positively coordinated. However, the various beliefs actually concern two distinct individuals. But then they are not co-referential and so this is not a case of manifest co-reference. The structure of Pam's beliefs is that of positive coordination without manifest co-reference (c.f. *Fig. 1*).<sup>9</sup>

We see, then, that it is possible for two mental representations to be positively coordinated without being manifestly co-referential. But if two mental representations are manifestly co-referential they are necessarily positively coordinated.

I take the fact that the framework allows us to draw a distinction between manifest and apparent co-reference to be a virtue of the proposed framework. It allows us to explain the difference between successful and unsuccessful (or lucky) action. A precondition for successful action is that the motivating belief/desire pair concerns the same individual.<sup>10</sup> To see why this is the case, let's say Pam forms a desire on Monday to meet *that* person, referring to twin<sub>1</sub>. On Tuesday she notices the mail carrier wearing salsa shoes and forms the belief that a way for her to meet *that* person, referring to twin<sub>2</sub>, is to go to the local salsa club. The only way for Pam to satisfy her desire is to meet twin<sub>1</sub>. However, twin<sub>2</sub> is the only one of the two who enjoys salsa, and so even though Pam's attitudes are positively coordinated, her going to the salsa club will result in an unsuccessful action since she does not get to satisfy her desire to meet twin<sub>1</sub>. We

---

<sup>9</sup> There are ways of spelling out the details of this case that make it seem likely that Pam is so confused that the relevant mental tags fail to refer altogether. Singular mental tags may only have one referent. If the scenario were spelt out in such a way that it makes it seem prima facie plausible that each of Pam's beliefs would refer to both twins, I would simply say that the relevant mental tags fail to refer. However, there seem to be ways of spelling out the details of the story, according to which it seems more likely that Pam forms specific beliefs each day and that each belief refers to either twin<sub>1</sub> or twin<sub>2</sub>, but not both.

<sup>10</sup> I assume again here a simple belief/desire account of agency (cf. Davidson 1963). Nothing hangs on this.

could, of course, imagine a scenario in which both twins enjoyed salsa dancing; in that case, going to the salsa club would actually result in Pam meeting twin<sub>1</sub>. This, however, is not successful action in the relevant sense, since it is merely lucky that twin<sub>1</sub> has the same hobby as her sister. In general, if a belief/desire pair is positively coordinated, the resulting behaviour will only be successful in the relevant sense (and not merely lucky) if the belief and desire are actually manifestly co-referential.

The broader notion of positive coordination is of interest for psychological generalizations, since it captures behavioural dispositions regardless of outcome. From the thinker's perspective, manifest co-reference is the same as apparent co-reference. To see this, imagine a different scenario in which Pam is right and there is actually just one mail carrier. In an important sense Pam's behaviour would be the same in both cases. Pam would have the same behavioural dispositions in the case of manifest co-reference as in the case of merely apparent co-reference: In both cases she would go to the salsa club in order to satisfy her desire. We see, then, that the suggested framework has the virtue of explaining behavioural dispositions in general as well as the distinction between successful and unsuccessful action by appeal to a minimal amount of entities – referentialist content together with pointer relations and mental tags is all we need.

I have presented the key claims of Vehicle Relationism. In the next subsection I will expound briefly on why pointers obtain or fail to obtain.<sup>11</sup> We shall see that the pointers are a result of our general capacity to keep track of entities.

### *2.3. Keeping Track of Objects*

Our cognitive systems function so as to keep track of objects over time and across different sensory modalities. For instance, if you see a white cup in front of you and at the same time touch that cup, you may end up with beliefs such as

---

<sup>11</sup> The account I will provide is a mere sketch. The issues I will discuss in section 2.3 open interesting avenues for future research.

THAT CUP IS WHITE and THAT CUP IS SOLID. In normal cases, the sameness in reference will be manifest to you and you may trade on identity and conclude that something that is white is solid. The reason why you may do so is that your sub-personal cognitive mechanisms package the information in such a way that each piece of information is treated as concerning the same object. Campbell (1987) makes a similar point:

It is true that cognitive skills of the thinker are in play [in cases like the one above], as he keeps track of the object from modality to modality. But these are not *conceptual* skills of the thinker: they do not have to do with his abilities in conceptual reasoning. [...] The cognitive skills in question here belong to a sub-personal level; they are part of the cognitive substratum that makes a conceptual life possible at all. (Campbell 1987, 283)

One's person-level conceptual skills depend partly on such sub-personal cognitive binding mechanisms. The fact that two token representations are treated as though they concern the same object at the conceptual level is a result of our more fundamental ability to keep track of objects in general.

As Campbell notes, parallel remarks can be made concerning our ability to track an object through time:

The sameness of the objects around one which one encounters from time to time is not an ordinary empirical hypothesis, established by investigation on the part of the subject, on the strength of which he takes a particular sequence of encounters with the same thing. And though keeping track of objects from moment to moment is certainly a cognitive skill, it belongs to a more rudimentary level than conceptual computation. (Ibid.)

Again, the conceptual capacities involved in trading on identity depend on the cognitive system's keeping track of the relevant object at a sub-personal level. If a subject forms two beliefs at different times, the subject will only be rational in

trading on identity if the two beliefs are partly a result of the tracking mechanism treating the information as if it concerned the same object.

We see, then, that our sub-personal cognitive capacity to keep track of objects is a prerequisite for our ability to trade on identity. What I suggest is that pointer relations are to be understood as a result of such sub-personal tracking mechanisms. Whenever the cognitive system binds different pieces of information together as though they concern the same individual, the resulting beliefs have constituents that stand in pointer relations.<sup>12</sup>

Such tracking mechanisms are, of course, not infallible. For instance, sometimes the tracking mechanism binds pieces of information together in such a way that they are treated as concerning the same object even though the pieces of information concern different objects. Imagine for instance looking at a snake on the ground, following it with your eyes, thinking you've been looking at the same snake the entire time. However, unbeknownst to you, the initial snake was swapped by a different indistinguishable snake at the blink of an eye. In this case, your cognitive system is working normally, but due to uncooperative external circumstances it binds the information together as though it concerned the same animal. The result is that you will form beliefs which constituents are related by pointers, but that actually concern distinct objects.<sup>13</sup>

Recall that, on the picture I have in mind, the reference of a singular representation (i.e. a mental tag) is determined by the relation the representation bears to the referent (as opposed to e.g. the referent satisfying a set of descriptions). In the case of the snake, then, the singular representations have different referents. However, since the relevant beliefs are related by pointers (given the bindings effected by sub-personal tracking mechanisms), you will have a rational disposition to trade on identity. This is a case of apparent co-reference, as defined in the previous section.

---

<sup>12</sup> Recanati (2012, e.g. 98) also appeals to such sub-personal binding mechanisms when accounting for information distribution in mental files.

<sup>13</sup> This case is similar to that of Pam being confused, in the previous section. The reason why Pam's beliefs are pointer related even though they refer to distinct individual is that her cognitive tracking mechanisms treat information from both of the twins as though it concerned the same individual.

The suggestion, then, is this: Pointer relations are (part of) the representational output of sub-personal tracking mechanisms. If the cognitive system binds two pieces of information together as a result of the workings of such tracking mechanisms, the result is that the subsequent attitudes stand in pointer relations. If the cognitive system does not bind information together in this way, there will be no pointer relations between the relevant attitudes. The key claim, then, is that nothing else is needed to account for coordination in thought. The pointer relations are *direct* outputs of sub-personal binding mechanisms, and do not reduce to sameness or difference of intrinsic properties of representations at the conceptual level.

In the next section I will show how the suggested framework provides solutions to some of the longstanding puzzles within the philosophy of mind.

### 3. Vehicle Relationism and Cognitive Significance

I started out by presenting one of the central puzzles within the philosophy of mind: If all there is to the content of singular representations is their referent, how can it be that co-referential representations may play distinct roles in cognition? This question, of course, goes back to Frege's (1982) puzzle of Hesperus and Phosphorus. The puzzle is this: In Ancient Babylonia people used the name 'Phosphorus' to denote the brightest star visible in the morning. They used the name 'Hesperus' to denote the brightest star visible in the evening. Unbeknownst to them, 'Hesperus' and 'Phosphorus' refer to the same object, namely the planet Venus. Even so, there seems to be an important difference between knowing that Hesperus is Hesperus and that Hesperus is Phosphorus: The former is trivial, while the latter involves new knowledge about the world.

Further, in the evening the Ancient Babylonians believed that Hesperus was visible and that Phosphorus was not visible. If there is nothing more to the content of thoughts than Millian/Russellian propositions, the Ancient Babylonians seem to have had conflicting attitudes towards the same proposition. Assuming they were rational, how can this be?



One of the key assumptions of Vehicle Relationism relevant to this puzzle is that thinking is a triadic relation. This means that thinkers are related to propositional content only indirectly, via representational vehicles. On this framework the cognitive role of mental representations is not merely a matter of semantic content. Hence, the Vehicle Relationist need not introduce a more complex theory of content in order to explain the cognitive role of mental representations. What accounts for the Ancient Babylonians being rational in holding contradictory beliefs is that the singular mental tags in the relevant beliefs were not related by pointers. Since they did not stand in pointer relations, the co-reference of the mental tags was not manifest. As long as the co-reference is not manifest it is rationally permissible for a subject to attribute contradictory properties to the same referent.

In the case of trivial versus informative identity judgements, the Vehicle Relationist appeals to the *cognitive effect* of the different beliefs. Cognitive effect is the result of taking a given belief as input to an already existing belief base (c.f. Fine 2007, 78—85). For the Ancient Babylonians, their belief base included various beliefs about the same object, but not all of these beliefs were connected by pointers. More precisely, they had two sets of beliefs each of which contained beliefs that were pointer related to the other beliefs in the set, but none of which were pointer related to the beliefs in the other set. If we take the belief HESPERUS IS HESPERUS as input to such a belief base this would have minimal effect. If instead we take the belief HESPERUS IS PHOSPHORUS as input, the cognitive effect will be significant. The singular vehicles of the input belief (i.e. HESPERUS and PHOSPHORUS) enter into pointer relations with one set of beliefs each. The identity then allows the thinker to combine the information in each set – something that was not warranted prior to the informative identity judgement – and as a result she may make new inferences. After having made the identity judgement the thinker is, for instance, in a position to infer that one and the same object is visible in the morning and in the evening.

We started out with the claim that given that the content of singular thought is transparent to thinkers, it seems plausible that whether or not two thoughts express the same singular proposition should be transparent to the thinker as

well. The Millian/Russellian would have to deny this. Someone can believe the same Millian/Russellian proposition twice without being aware of this fact. When the Ancient Babylonians believed that Hesperus is a star and also that Phosphorus is a star, they believed the proposition <Venus, being a star> twice without being aware that the two thoughts concerned the same object. Given that sameness of Millian/Russellian proposition is not transparent to the thinker, the classic Millian/Russellian is in a good position when it comes to explaining how thinkers may fail to recognize sameness. However, it is clear that we have cases in which the thinker does recognize sameness of reference, and does so in a systematic way. We will see that the pointer picture offers an account of such systematicity within a Millian/Russellian framework.

### 3.1. *The Systematicity of Recognition*

The fact that recognition of sameness of propositional content is systematic becomes particularly clear in cases where two pairs of beliefs express the exact same pair of Millian/Russellian propositions, but where which belief pair a given individual entertains makes a difference to her behavioural dispositions. Consider for instance Lex Luthor wanting to kill Superman. He believes that a way to kill Superman is to shoot him with his nuclear kryptonite ray gun. Carrying his ray gun in his suitcase he sees Superman in front of him, but does nothing. How can this be? When seeing Superman in front of him he forms a belief that has the singular proposition <Superman, being in front of me> as content.<sup>14</sup> Still, this belief does not interact with his other beliefs in such a way that he will choose to shoot the guy in front of him. Why is this?

Lex Luthor has two distinct sets of singular beliefs concerning Superman. Some of these express the propositions <Superman, wearing a red cape>, <Superman, wearing blue tights>, <Superman, being called 'Superman'> and so on. At the

---

<sup>14</sup> In the case of 'Superman' and 'Clark Kent' we do not have a neutral way of talking about the referent. In the case of 'Hesperus' and 'Phosphorus' we can use 'Venus' to signal neutrality, but we do not have anything similar in the case at hand. Hence, whenever 'Superman' is used to indicate the individual as a constituent of a proposition I intend the use to be neutral between beliefs representing him as *Superman* or as *Clark Kent*.

same time, he has a set of beliefs concerning Superman that express propositions such as <Superman, wearing glasses>, <Superman, wearing grey suits>, <Superman, being called 'Clark Kent'> and so on. The person Luthor sees in front of him fits the latter set of descriptions. As a result, Luthor's new belief that expresses the proposition <Superman, being in front of me> will enter into pointer relations with the beliefs in the latter set and not with the beliefs in the set consisting of beliefs about Superman being dressed in a cape and tights.

At the same time, Luthor's desire to kill Superman and also his belief that a way to kill Superman is to shoot him with his ray gun stand in pointer relations to the set of beliefs taking Superman to be wearing the superhero costume. The belief/desire pair is, however, not pointer related to the set of beliefs that his newly formed belief is pointer related to. This explains why, in a systematic way, Luthor only acts on his desire to kill Superman under certain circumstances, namely the circumstances in which the person he sees fits the relevant beliefs in the set. Importantly, what set Luthor's newly formed belief is pointer related to is a result of sub-personal cognitive binding mechanisms (c.f. section 2.3. above).

We see that which pointer relations a given representation enters into depends, at least in some cases, on the thinker's already existing beliefs about the referent. Importantly, however, this is very different from the Fregean account of senses construed as definite descriptions. On the pointer picture there is no abstract public entity that one grasps in virtue of acquiring a certain concept. Nor is the referent of mental tags determined by such associations. An individual may have false beliefs about a given individual and still successfully have a singular thought about that individual.<sup>15</sup> Similarly, a single individual may associate the exact same properties with two distinct individuals and still successfully refer to each one of them individually. According to the particular version of Vehicle Relationism I have in mind, the referent of a singular thought is not picked out via the descriptions the thinker associates with the referent. However, the associated descriptions (partly) explain the way information is stored in

---

<sup>15</sup> The suggested framework thus resembles the mental file framework (discussed in Chapter 1) in that the reference of a singular thought is not established satisfactorily (c.f. Recanati 2012, 57).

cognition. Importantly, these descriptions are not part of the semantic content of the singular representations referring to the object in question. Wearing certain clothes is not part of the content of Luthor's mental tags referring to Superman – the content of singular mental tags are just their referents.

In order to explain systematicity in behavioural dispositions, then, we cannot just appeal to the propositional content alone. As we saw in section 2.2, psychological generalizations also depend on pointer relations. In order to explain and predict how an individual will behave we need to take the pointer relations into account. A related question concerns the relation between an individual's propositional attitudes and her utterances. If an individual has two distinct beliefs expressing the same propositional content, how can it be that an assertion expressing the very same proposition may express only one of these beliefs? In what follows I will give an account of this in terms of pointer relations.

### *3.2. The Relation between Propositional Attitudes and Utterances*

Given that the ancient Babylonians had two distinct beliefs expressing the proposition <Venus, being a star>, how can it be that only one of these beliefs was expressed by their utterances of the sentence "Hesperus is a star"? Here I take the explanation to be partly metalinguistic. The belief HESPERUS IS A STAR is pointer related to a belief that expresses the proposition <Venus, being called 'Hesperus'>. In contrast, the belief PHOSPHORUS IS A STAR is pointer related to a belief expressing the proposition <Venus, being called 'Phosphorus'>. This, of course, is not to say that every time someone expresses a belief she has to consciously access such metalinguistic beliefs. Rather, the mere existence of such beliefs explains our general ability to express our attitudes linguistically. These beliefs are available to our cognitive systems (e.g. the language faculty) even though individuals may not be consciously aware that such beliefs are exploited.

But, now, what are we to say about Peter from Kripke's (1979) puzzle of Paderewski? According to the story, Peter encounters the same person, Paderewski, on two separate occasions without recognizing him as the same. He

forms various beliefs about Paderewski, but only some of the beliefs are recognized by Peter as concerning the same individual. Peter recognizes that all of the beliefs concerning Paderewski formed on their first encounter concern the same individual. Similarly, he recognises that all beliefs formed on the basis of their second encounter concern the same individual. But he does not know that the beliefs formed on the first encounter concern the same person as the beliefs formed on the basis of the second encounter. On the first day, Peter forms the belief that Paderewski has musical talent. On the second day, he forms the belief that Paderewski does not have musical talent. According to the story, Peter has good reasons to accept both of these claims, and so he is rational in holding seemingly contradictory beliefs. How can this be?

This puzzle creates problems for Frege. Frege (e.g. 1892) famously appeals to a distinction between sense and reference, where senses account for coordination. As we have seen, sameness or difference of reference may come apart from sameness or difference in the cognitive role of concepts. At the same time, two concepts that are positively coordinated always play the same role in cognition. Likewise, negatively coordinated concepts will not play the same cognitive role in this way. If senses are to explain coordination, then, sameness or difference of sense cannot come apart from coordination. Frege made a distinction between subjective associated descriptions (what he calls *ideas* (1892)) and senses. The latter are taken to be public and sharable. The puzzle of Paderewski shows that sameness of sense does not guarantee that two concepts will play the same role in cognition, since there is only one public sense associated with Paderewski. If this is the case, senses cannot account for coordination.

While this creates problems for Frege due to there only being one public sense associated with Paderewski, the Vehicle Relationist explanation of how Peter may be rational despite believing contradictory things about the same person is the same as in the puzzle of Hesperus and Phosphorus: Peter's beliefs about Paderewski are not connected by pointers. Someone can be rational in believing contradictory propositions as long as the contradiction is not manifest to the thinker. In this particular case, the contradiction is manifest only as long as the co-reference of the two tokens of PADEREWSKI is manifest. The co-reference is

not manifest to Peter, since the relevant mental tags are not connected by pointers. In order to explain the relation between Peter's beliefs and his assertions, however, we need to give further specifications of the general relation between attitudes and utterances.

When it comes to explaining the relation between Peter's beliefs and his assertions, there are two questions: First, how can it be that Peter may use the same English sentence to express two distinct beliefs? To be clear, the puzzle is not how anyone ever can use the same sentence to express different beliefs – it is obvious that the same sentence may be uttered twice to express different beliefs due to a difference in context. Rather, the puzzle concerns the fact that it seems possible for Peter to express two distinct beliefs by the same sentence in the same context. Consider Peter uttering: "Paderewski is a human being". This sentence expresses the proposition <Paderewski, being a human being>, a proposition which Peter happens to believe twice, in two different ways. How can it be that when Peter utters this sentence he only expresses one of these beliefs? Further, how do we determine which of the two beliefs he expresses on a given occasion?

Just as in the case of Hesperus and Phosphorus, Peter has some metalinguistic beliefs. He believes the propositions <Paderewski, being called 'Paderewski'>. This time, however, the case is slightly different, since Peter's relevant metalinguistic beliefs express the very same proposition. Even though Peter takes there to be two public names 'Paderewski' this is not actually the case. The two tokens of the name 'Paderewski' in the metalinguistic beliefs actually refer to the same public name. How, then, do we explain this?

According to Vehicle Relationism, two beliefs that express the same proposition may nonetheless be distinct beliefs. Synchronic identity, i.e. identity at a given time, of propositional attitudes requires sameness of content plus identity of pointer relations. Since Peter's two metalinguistic beliefs are pointer related to different attitudes they are distinct beliefs.

When Peter utters, “Paderewski is a human being”, his cognitive system takes a certain route: Very crudely, it starts at one of Peter’s two beliefs expressing <Paderewski, being a human being>, and goes either directly or indirectly via the metalinguistic belief that is pointer related to the belief he intends to express. Even though both routes from Peter’s two beliefs have the result of Peter expressing the same English sentence with the same propositional content, the cognitive route causing the utterance is different. This is why, even though Peter’s utterances mean the same regardless of which belief he expresses, Peter only expresses one of these beliefs by any given utterance of the relevant sentence.

Note that the pair of beliefs at each stage has the same propositional content: First, both of Peter’s beliefs concerning Paderewski being a human being express the same proposition. Second, both the metalinguistic beliefs express the same proposition due to there being only one linguistic term ‘Paderewski’. Finally, the resulting utterance will express the same proposition regardless of which of the two beliefs Peter intends to express. The representational vehicles are thus explanatorily essential to any theory that takes content to be classic Millian/Russellian propositions.

I have shown how one may account for some of the central difficulties facing Millian/Russellian accounts of mental content. If we take thinking to be a triadic relation in which the thinker stands in a direct relation to representational vehicles and in an indirect relation to propositional content via such vehicles, we may account for the cognitive role of thoughts in terms of relational features of such vehicles (i.e. whether or not two vehicles stand in pointer relations) rather than by a direct appeal to the content of thoughts.

In what follows I offer a comparison between Vehicle Relationism and two views that in many ways resemble the proposed framework, namely Heck’s (2012) Formal Relationism and Fodor’s (1975, 2008) Language of Thought Hypothesis (LOTH). This will help make the proposed framework clearer.

#### 4. Comparison with other Views

Although the view I propose in this paper offers a new account of coordination in thought, it bears some similarities to certain well-known theories. In what follows, I will point to similarities and differences between Vehicle Relationism and Heck's (2012) Formal Relationism as well as Fodor's (1975, 2008) LOTH. We may distinguish between two distinct although highly related questions in connection to coordination relations: (i) what distinguishes two thoughts that express the same Millian/Russellian proposition but that nonetheless play distinct roles in cognition? and (ii) in virtue of what do coordination relations in thought obtain? Heck only thoroughly addresses the first question, but they allude to what they suspect might be a correct response to the second question, namely that coordination relations hold in virtue of a Fodor style language of thought. As will become clear, I agree with Heck's response to the first question, but I will argue that if combined with Fodor style LOTH the framework is not a genuine relationist account as defined at the outset. By 'genuine relationism' I mean views that give relationist answers to both questions. I will then spell out the alternative response to question (ii) that follows from the pointer picture discussed above. The suggested view is a genuine relationist account of coordination relations.

#### *4.2. Comparison with Heck's Formal Relationism*

In their (2012), Heck develops the framework of *Formal Relationism*. They argue that the key to solving Frege's puzzle for thought is not to look at the content of thoughts, but rather to look at how the relevant thoughts are formally related to each other. Such relations, they argue, need not supervene on sameness or difference in content.

Heck considers the difference in the cognitive role of (1) and (2):

- 1) SAMUEL CLEMENS HAS DIED
- 2) MARK TWAIN HAS DIED



CLEMENS and TWAIN refer to the same individual, but as we have seen, this is no guarantee that they play the same role in cognition. Heck illustrates this by considering a case involving an individual, Fred, who when he comes to believe (1) gets devastated, whereas he does not get equally saddened when coming to believe (2). Heck argues that

1. What distinguishes the belief that Clemens has died from the belief that Twain has died is nothing intensional. In particular, these beliefs have the same content.
2. If we are to be able to explain Fred's behavior in cognitive terms, there must be some difference between these beliefs that plays a role in psychological explanation. But no *intrinsic* difference between these beliefs plays that role. The explanatorily relevant difference is an extrinsic, relational one. It concerns how these beliefs are related to other of Fred's beliefs. (Heck 2012,144)

The basic claim is this: Frege cases show that inferential patterns of beliefs are partly explained in terms of whether or not they stand in a certain formal relation. Such formal relations are similar to the ones found in formal logic in inferences where sameness of reference is presupposed rather than stated explicitly:

Saying that Fred's belief that Clemens has died is 'formally related' to his belief that Clemens is his neighbor means that the beliefs have the feature we aim to capture in formal logic when we represent them this way:  $D(c)$ ,  $c = \text{the } N$ ; rather than this way:  $D(t)$ ,  $c = \text{the } N$ . (Ibid., 145)

This, they argue, is all that is really established by Frege cases. It does not follow directly from Frege's puzzle that there must be a difference in content of CLEMENS and TWAIN. Rather, what follows from such Frege cases is that there must be a relational difference between the pairs [CLEMENS, CLEMENS] and [CLEMENS, TWAIN].

I agree with Heck that the relational aspects of mental representations are the key to understanding cognitive significance and that nothing about content follows directly from Frege's observation. In fact, I take this to be one of the key claims of Vehicle Relationism.

Heck's main focus is on the consequences of Frege cases for intentional explanation and in stating intentional laws. They argue that for this specific purpose we only need the formal relations and as such, "the sorts of formal relations among beliefs that must be mentioned may be treated as psychologically primitive: We can make reference *directly* to these relations in giving intentional explanations and in stating intentional laws" (Heck 2012, 157). As a result, psychological generalizations only require Millian/Russellian propositions and direct reference to the relations between thoughts. This takes us back to the discussion in section 2 above. Behavioural dispositions do not only depend on the propositional content of desires and beliefs, but also on whether or not the desires and beliefs are positively coordinated. We may make a more general relationist claim based on Heck's argument:

**General Relationist Claim:** All that is needed to account for Frege cases are relational aspects of thought.

All relationists are committed to this general claim. However, as mentioned, there is a further question concerning coordination relations: In virtue of what do such relations obtain? Heck sets this question aside, but they suggest that "it may well be that two beliefs' standing in such a relation is, as a matter of fact, ultimately to be explained in terms of facts about how cognitive states are implemented" (Heck 2012, 157).

In this regard, the framework in terms of pointer relations suggested in this paper goes one step further than Heck's formal relationism: Heck only aims at answering the first question, about what it is that distinguishes two thoughts that express the same Millian/Russellian proposition but that nonetheless play distinct roles in cognition. In contrast, in addition to answering the first question, the Vehicle Relationist offers a response to the second question of in virtue of

what coordination relations in thoughts obtain. I will elaborate on this in section 4.1. below.

Heck makes clear that they do not want to deny Fodor style LOTH:

It would be strange to regard the ‘formal relations’ as brute. So their obtaining or failing to obtain must supervene on something else. But it is no part of my view here to say what that is. It might be a lot of things. For what it’s worth, I suspect that the language of thought hypothesis is true and that formal relations supervene on Mentalese syntax. (Heck 2012, 159)

I will argue that if combined with Fodor style LOTH, Formal Relationism becomes a version of Intrinsicism, as defined at the outset, rather than genuine Relationism about coordination in thought. In order to see why this is the case, let’s take a closer look at Fodor’s framework and how it relates to Vehicle Relationism.

#### *4.1. Comparison with Fodor’s Language of Thought Hypothesis*

The key claim of Fodor’s LOTH is that thinking takes place in a mental language, usually just referred to as ‘Mentalese’. This language is much like natural languages in that it is made up of syntactic entities that compose into sentence-like structures:

This language consists of a system of representations that is physically realized in the brain of thinkers and has a combinatorial syntax (and semantics) such that operations on representations are causally sensitive only to the syntactic properties of representations. (Aydede 2010)

On this view, then, thoughts are mental representations that have a compositional semantics. Fodor thus accepts a version of the representational

theory of mind (RTM), which is the hypothesis that propositional attitudes consist of relations between subjects and mental representations:

RTM is a claim about the metaphysics of cognitive mental states and processes: Tokens of cognitive mental states are tokens of relations between creatures and their mental representations. (Fodor 2012, 5)

Further, he says that

according to RTM, Mentalese singular terms, predicates, and the like refer to things in the world and [...] expressions of Mentalese are the representations over which mental processes are defined. (Fodor 2008, 93)

According to this view, mental representations stand in relations to things in the world, but the cognitive system is only directly sensitive to the syntactic features of thought, and computes on syntactically specified mental representations. In this way, thinking becomes a triadic relation between thinker, mental representations and propositional content; thinkers are related to propositional content via representational vehicles.

It becomes clear that Vehicle Relationism bears many similarities to Fodor's LOTH. Both frameworks take thinking to be a triadic relation between thinkers, representational vehicles and propositions. Both theories hold that cognitive processes are only sensitive to syntactic features of thoughts (i.e. representational vehicles). Both frameworks thus take the explanation of coordination to essentially be a matter of syntactic features: Co-referential mental representations may differ only in properties found at the level of representational vehicles.

The main difference between Fodor's LOTH and Vehicles Relationism is how the frameworks construe the syntax of the language of thought. Fodor takes mental vehicles to be highly similar to words in natural language: They can be construed

as having syntactic *forms* that are recognized by the cognitive system.<sup>16</sup> Two vehicle tokens that have the same syntactic form are of the same type and so they play the same role in cognition. In other words, on this view coordination is explained in terms of sameness or difference in such syntactic forms. This provides an answer to the second question concerning in virtue of what coordination relations obtain (or fail to obtain): Two vehicles that have the same form are positively coordinated, while two vehicles that have distinct forms are negatively coordinated.

We see then that Fodor's view is an intrinsicist account of coordination: Coordination is accounted for in terms of intrinsic representational features, namely vehicles being of a certain type in virtue of having a certain syntactic form.<sup>17</sup> This is a good point at which to return to Heck's inclination to think that "the language of thought hypothesis is true and that formal relations supervene on Mentalese syntax" (Heck 2012, 159). If this is correct and Mentalese syntax is construed the way Fodor suggests, Formal Relationism is not really a relationist account of coordination. This is because coordination relations, on this view, ultimately are accounted for in terms of intrinsic representational features of thoughts and thought constituents. On this picture it is the case that, although it might be correct that only the formal relations are needed for psychological generalizations, such relations could ultimately be reduced to intrinsic representational features of thoughts. Whether or not Formal Relationism is a genuine relationist framework thus depends on what theory of implementation one adopts.

Vehicle Relationism offers an alternative story of how the coordination relations

---

<sup>16</sup> Note that the talk about *forms* is of course purely metaphorical: type identity of primitive Mentalese concepts is not determined by their physical 'shape' in the brain, since the same concept may be physically realized in many different ways in the head of different thinkers, or even the same thinker at different times. The metaphor comes from Fodor (2008, 80) where he illustrates sameness or difference in primitive concept types by appeal to how we determine sameness of types of letters: "We distinguish 'dog' tokens from 'cat' tokens by their spelling, but we don't distinguish 'a' tokens from 'b' tokens that way, since 'a' and 'b' don't, of course, *have* spellings. What they have is *shapes*; and their shapes are different in ways to which our visual system is responsive; if they weren't, we wouldn't be able to read. Likewise *mutatis mutandis* for the way the minds draw type distinctions between tokens of basic mental representations".

<sup>17</sup> Again, I base the notion of an intrinsic representational feature on Gray (2017). As mentioned previously, the relevant features are those that can be stated without reference to another representation. See the Introduction chapter, section 3.1, for elaboration on this notion.

are implemented. According to Vehicle Relationism, we do not need to appeal to syntactic forms or the like in order to account for coordination in thought. Instead, the syntax of the language of thought must be supplemented with pointer relations that hold between mental tags. When it comes to coordination, the cognitive system is sensitive to such relations rather than intrinsic syntactic features of the vehicles. Representational vehicles construed as mental tags do not have distinct syntactic forms. At the level of vehicles, all mental tags have the same intrinsic features. Tags only differ to the extent that they stand in different pointer relations or express different semantic contents. If two thoughts are positively coordinated the cognitive system “reads off” this relation directly, rather than treating the representations as the same due to their intrinsic features. According to this view, the pointer relations are representationally primitive.<sup>18</sup>

The formal relationist who takes coordination relations to supervene on the syntax of the language of thought thus has a choice between on the one hand a classical Fodor style construal of the language of thought, which renders the theory an Intrinsicist account of coordination relations, and on the other, Vehicle Relationism which takes the language of thought to be irreducibly relational. Both alternatives are compatible with the General Relationist Claim, but only the latter is compatible with genuine relationism about coordination relations.

A natural question to ask at this point is whether or not we have any reason to think that the language of thought is irreducibly relational, except from simply presupposing relationism about coordination. One such reason can be drawn from Pinillos’ (2011) argument that coordination is a non-transitive relation. Consider the following thought:

---

<sup>18</sup> By this I mean that the pointers cannot be reduced to sameness or difference in other features of mental representations. That, of course, is not to say that pointer relations cannot be reduced to sameness or difference in *any* property. I have not said anything about how the pointer picture may be neurologically implemented, and make no claim about this matter in this paper. It might turn out that the pointers ultimately obtain (or fail to obtain) as a result of sameness or difference in neurological firings.

- 1) HESPERUS<sub>1</sub> IS PHOSPHORUS<sub>2</sub> AFTER ALL, SO HESPERUS-SLASH-PHOSPHORUS<sub>1,2</sub> MUST BE A VERY RICH PLANET.

In this case, each of the simple representations (i.e. HESPERUS and PHOSPHORUS) are positively coordinated with the *slash-concept* (i.e. HESPERUS-SLASH-PHOSPHORUS), but at the same time the simple representations are negatively coordinated with each other. Such slash-concepts are the result of informative identity judgements of the form  $a = b$ . When someone learns that Hesperus is Phosphorus they will in normal circumstances form a new representation of the form seen in (1). This new concept is the one that will normally figure in thoughts after the individual has made such identity judgements. Even so, the individual is still able to use the simple representations, as seen in (1). If this were not the case, people could not rationally think thoughts such as I USED TO THINK THAT A WAS NOT B, after having formed the belief  $A = B$ .

If this is correct, coordination relations are non-transitive.<sup>19</sup> However, if coordination relations supervene on Fodor style LOT, the coordination relation would be transitive since identity of syntactic form is a transitive relation. If, in contrast, coordination relations supervene on the language of thought, relationally construed, we face no problem accounting for Pinillos' observation. There's nothing about the nature of pointer relations that gives us reason to think that they are transitive. The possibility of coordination relations being non-transitive gives us at least one reason to prefer Vehicle Relationism to the traditional LOT.

To sum up: I agree with Heck that all that is needed for psychological explanation is content, understood as Millian/Russellian propositions, as well as relational features of thought. I disagree, however, with Heck when it comes to the

---

<sup>19</sup> Soames (2010) also suggests that coordination is a non-transitive relation. His example is (26): "John fooled Mary into thinking he wasn't John". He suggests that we "revise the notions of coordination and representing *as the same* so that the term-occurrences in the complement of (26) can be coordinated with the subject of 'fooled' without being coordinated with each other" (Soames 2010, 474 n.1).

implementation of such relations. According to Vehicle Relationism, coordination relations hold in virtue of primitive relations between representational vehicles rather than in virtue of representational vehicles being of the same type. Only if we combine Formal Relationism with a relational framework of how coordination relations are implemented will the framework be a genuine relationist account of coordination.

In the next section I will compare Vehicle Relationism to Fine's (2007) Semantic Relationism. Vehicle Relationism combines some elements of Fodor style LOTH with elements of Semantic Relationism. Like the Fodorian LOTH it holds that thinking is a triadic relation, and like Semantic Relationism it takes coordination to be explained in terms of relational properties of mental states rather than intrinsic features. However, the proposed framework differs significantly from both of these other accounts of coordination. Ultimately I will argue that Vehicle Relationism is preferable to Semantic Relationism.

### 5. Vehicle Relationism vs. Semantic Relationism

Thus far I have spelt out the central details of Vehicle Relationism: Coordination relations are construed as pointer relations that have mental tags as relata. The positive account is compatible with a highly minimal classical Millian/Russellian view of propositional content. Coordination obtains at the level of representational vehicles and manifest co-reference is simply a matter of coordination plus sameness of reference. The pointer picture combined with classical Millian/Russellianism about propositional content is thus a highly attractive framework due to its explanatory power and parsimony. In what follows, I will point to similarities and differences between this framework and Semantic Relationism.



### 5.1. Comparison with Semantic Relationism

Fine's (2007) Semantic Relationism is first and foremost a theory of coordination in language. Even so, he expands his view and argues that

just as there are semantic relationships between expressions that are not to be understood in terms of their intrinsic semantic features, so there are representational relationships between the constituents of thought that are not to be understood in terms of intrinsic representational features. [...] If this is right, then it means that the apparatus of coordinated content is equally applicable within the realm of thought (Fine 2007, 66).

Further, he says that

for a thought or thoughts to represent an object as the same is for it to be a representational requirement that the object of the thought or thoughts is the same. [...] the intentionality of thought will be given by a body of representational requirements, which indicate how our various thoughts represent what they do; and coordination will be achieved in either case when those requirements demand an identity in what the language of thought is about (Ibid., 72).

To say that a thought represents an object as the same, as opposed to representing the object *to be* the same, is just to say that the relevant representations are positively coordinated. Thus far, the general idea is that there are representational relationships that are to be taken as primitives rather than reduced to sameness of intrinsic representational features. I agree with this general idea.

As mentioned, the main difference between Semantic Relationism and Vehicle Relationism is what the views say about the nature of coordination relations. We may ask: what are the relata of the coordination relations? The Vehicle Relationist holds that the relations are realized at the level of representational vehicles: There are primitive relations holding directly between such vehicles. As

I have shown, this view is compatible with a classical Millian/Russellian view of propositions. In contrast, the Semantic Relationist takes the relata of coordination relations to be constituents of propositions. Unlike the Vehicle Relationist, then, the Semantic Relationist is committed to a relational account of propositions:

The content of a belief will be given by a coordinated rather than by an uncoordinated proposition. Thus we may distinguish between the content of the belief that Cicero is Tully (where this is the negatively coordinated proposition) from the content of the belief that Cicero is Cicero (where this is the positively coordinated proposition). This is already a great advantage on the usual referentialist view, which is unable to make any such distinction without either distorting the logical form or appealing to some notion of sense or “guises” (Fine 2007, 77).

In section 1, I mentioned that the Semantic Relationist approach to coordination is similar to the Fregean approach in that the Semantic Relationist takes cognitive difference to imply a difference in semantic content. That is, on this view sameness of propositional content is transparent to thinkers. Even though the Semantic Relationist denies that reference is all there is to the semantic content of thoughts the account is supposed to be Millian/Russellian in nature (c.f. Fine 2007, 53-54). According to Semantic Relationists the semantic content of HESPERUS is just its referent. Likewise the semantic content of PHOSPHORUS is also its referent. Hence, the two concepts have the same intrinsic semantic properties. Even so, the pairs [HESPERUS, HESPERUS] and [HESPERUS, PHOSPHORUS] differ in their overall semantic properties as a result of the first pair being positively coordinated while the latter is negatively coordinated. Semantic Relationists thus hold that the thoughts HESPERUS IS HESPERUS and HESPERUS IS PHOSPHORUS differ in their overall semantic content, and this is why the two thoughts may play different roles in cognition. Note that by making this claim, the semantic relationist rejects compositionality: Two thoughts may have the same structure and contain constituent concepts that are true synonyms but

still fail to express the same propositional content, as in the case of the thoughts HESPERUS IS HESPERUS and HESPERUS IS PHOSPHORUS.

Note the last sentence of the quote above, where Fine says that Semantic Relationism does not need to appeal to the notion of sense or “guises”. In the next section, we shall see that this is not actually the case. In what follows I provide arguments to the effect that Vehicle Relationism provides a better account of coordination than Semantic Relationism. The first argument concerns the difference in complexity of propositional content. Semantic Relationism ends up having to postulate further levels of semantic content in addition to classical Millian/Russellian propositions. This is where the Semantic Relationist seems committed to something akin to “guises”. The second argument concerns the nature of the semantic relations and how they may account for cognitive effect. I argue that in order to account for the cognitive effect of coordination relations the semantic relationist must postulate something akin to pointer relations. If the semantic relationist needs pointers and the pointer picture can do all explanatory work without having to invoke an unnecessary complex semantics, the pointer picture – and thereby Vehicle Relationism – would be the better relationist alternative.

### *5.2. Semantic Relationism: Propositional content*

According to the semantic relationist, the difference between the two sentences “Hesperus is Hesperus” and “Hesperus is Phosphorus” is that the former expresses a positively coordinated proposition whereas the latter expresses a negatively coordinated proposition. But what about sentences that contain only one objectual component, such as “Hesperus is a planet” and “Phosphorus is a planet”? In such cases, there are no two concepts that may be coordinated. Hence, the sentences seem to express the very same proposition and have the same truth conditions.

In his (2010a) Fine develops the framework of Semantic Relationism further as a response to criticism made by Soames (2010). Soames criticises Fine’s account of

the semantics of belief reports. Given that two sentences such as “Hesperus is a planet” and “Phosphorus is a planet” express the same proposition on Fine’s view, what is the Semantic Relationist to say about belief reports such as “John believes that Hesperus is a planet” and “John believes that Phosphorus is a planet”? It seems that the two reports will express the same proposition, but at the same time it appears that one may be true while the other may be false. The focus of this paper is not on the semantics of language, including belief reports. I will therefore put aside this specific worry for now. However, I will raise a problem about thoughts that I think forces the semantic relationist to give a response similar to the one given as a response to Soames’ objection. In what follows I present the objection. I then present the further developments of Semantic Relationism that Fine makes as a response to Soames’ objection and show how this also provides a response to my objection. Finally, I argue that the resulting Semantic Relationist account becomes extremely metaphysically complex and as a result it becomes an unattractive alternative for those who are sympathetic to referentialism.

### *5.2.2. The Objection*

We saw that, for the Semantic Relationist, the thoughts HESPERUS IS HESPERUS and HESPERUS IS PHOSPHORUS differ in their overall semantic content as a result of the former expressing a proposition in which the objectual constituents are positively coordinated whereas the latter expresses a proposition in which the objectual constituents are negatively coordinated. We also saw that two structurally identical thoughts that contain concept tokens with the same intrinsic semantic features and that do not contain any coordination relations have the same overall semantic value. This renders the thoughts HESPERUS IS A PLANET and PHOSPHORUS IS A PLANET semantically identical. But, then, the question is this: If the Ancient Babylonians had the belief that Hesperus is a planet and then formed the belief that Phosphorus is a planet, how can the latter be a new belief given that they already believed that very proposition? That is, we see why someone who already knows that Hesperus is Hesperus would form

a new belief when informed that Hesperus is Phosphorus; the overall semantic properties of the proposition already believed differs from the overall semantic properties of the proposition that is the content of the newly formed belief. But in the case at hand, the proposition believed, namely that Hesperus is a planet has the same overall semantic properties as the content of the newly formed belief that Phosphorus is a planet.

Further, what can the Semantic Relationist say about someone believing and denying the same proposition in cases such as someone thinking that Hesperus is visible but that Phosphorus is not visible? How can someone at the same time both believe and deny this proposition, assuming that sameness or difference in semantic content is transparent to thinkers? There is nothing more to Fine's framework of coordination than propositions. On his account thinking is a dyadic relation holding between thinkers and propositions.<sup>20</sup> What, then, is going on when someone believes and denies the same proposition or when someone gains new information despite already believing that very proposition?

As far as I can see, the only way for the Semantic Relationists to respond to this worry is for them to make further developments of their framework along the lines of Fine (2010a). In what follows I will spell out the details of this proposal and explain why this would provide a solution to the worries raised above.

### 5.2.3. *Fine's Reply*

As mentioned, Fine develops his framework further specifically as a response to Soames' criticism pertaining to the semantics of belief reports. However, as we shall see, these developments also provide a solution to the worry raised above. Let's take a look at Fine's reply.

---

<sup>20</sup> Fine does not say much about the metaphysics of thoughts. What's important is that he does not give any explanatory import to the notion of representational vehicles or the like. He holds that "thoughts do not appear to have the same kind of clear syntax as sentences. [...] this provides, by the way, yet another reason not to think of coordination syntactically in terms of the repeated use of the same symbol" (Fine 2007, 73).

Fine (2007) makes a distinction between uncoordinated and coordinated propositions. The former amount to what he calls *primary* content and are just classical Millian/Russellian propositions. Coordinated propositions are what he calls *secondary* content, and these are Millian/Russellian plus coordination relations. This means that the beliefs that Cicero is Cicero and that Cicero is Tully have the same primary propositional content, but since they differ in coordination relations they express different secondary propositional contents. We see that the primary propositions are multiplied at the secondary level. The two propositions that Cicero is Cicero and that Cicero is Tully are the same at the primary level, but in order to make sense of the coordination relations they must be distinct at the secondary level. However, the thoughts HESPERUS IS VISIBLE and PHOSPHORUS IS VISIBLE express the same primary and secondary content since there are no two occurrences of individuals to be coordinated.

Fine (2010a) introduces a further layer of content, namely *tertiary* content. The tertiary content is the universal body of coordinated propositions. Here we find all the various secondary propositions:

Consider all of the propositions that are realized in propositional acts such as believing, asserting, intending and the like. They form a vast coordinated body of propositions, where some of the occurrences of any given individual in these propositions are linked to others and some are not. (Fine 2010a, 479)

At this level, the coordinated propositions do not only have their intrinsic coordination relations (i.e. coordination relations within a single proposition), they also enter into coordination relations with each other. For instance, the proposition that Cicero is Cicero will have two coordination links to the first relatum of the proposition that Cicero is Tully.

Further, at the level of tertiary content, there will be more than one occurrence of a single secondary proposition. That is, some secondary propositions are further multiplied at the tertiary level. For instance, since the beliefs HESPERUS IS A PLANET and PHOSPHORUS IS A PLANET involve the same first and secondary

proposition, but nonetheless enter into different coordination relations at the tertiary level, there must be two distinct, although intrinsically identical, propositions in the universal body of coordinated propositions.

Fine calls the propositions at the tertiary level of content *token propositions*, and says that “the identity of a token proposition is partly given by its intrinsic content, i.e. by the underlying coordinated proposition, and partly by its coordinative links to other token propositions” (Fine 2010a, 479). In order to posit *relata* for the coordination relations at the tertiary level Fine also introduces the notion of *token individuals*, which are the objectual constituents of token propositions. He says that “token individuals are a little like ‘guises’ or individual concepts, but there is no special descriptive content or mode of presentation with which they must be associated” (Ibid., 480).

One thing to note at this point is that Fine explicitly appeals to something that is, as he says, a little like ‘guises’. Although he rejects the Fregean idea that such guises have descriptive contents, these guises seem to be, in one way or another, ways of representing the referent. Even though there may be no intrinsic semantic difference between such guises that explain how they may play different roles in cognition – there may be only relational differences – it is still a fact that the semantic relationist needs this further level of semantics in addition to the Millian/Russellian propositions in order to give a full account of cognitive significance. It is only once we have this further level of semantics that we may posit a response to the objection raised in the previous section: The reason why it is one thing to think that Hesperus is visible and another to think that Phosphorus is visible is that these propositions are different token propositions, and that is just to say that the propositions enter into different coordination relations in the universal body of propositions.

This is a good point at which to highlight a central difference between Semantic Relationism and Vehicle Relationism. The Vehicle Relationist would explain the difference between the two beliefs that Hesperus is visible and that Phosphorus is visible by saying that the representational vehicles which constitute the two thoughts are not related by pointer relations. Further, the two thoughts would

enter into different pointer relations if the thinker has more propositional attitudes concerning Venus. The Vehicle Relationist does not need to posit a semantic difference between the two thoughts; according to this view, coordination in thought is not explained in terms of propositional content.

In contrast, Fine needs to introduce multiple tokens of the same Millian/Russellian proposition in order to account for the different coordination relations it may enter into. Fine's 'guises' cannot be understood as representational vehicles; they are semantic in nature. For every possible combination of coordination relations, Fine must say that there are distinct token propositions and token individuals. Given the vast variety of how people may fail to know that concept tokens are co-referential, this creates a vast body of token propositions. Take for instance Peter, from Kripke's Paderewski case. His belief that Paderewski is a politician would create a new token of that proposition, since the proposition believed by Peter is not coordinated with the token proposition that Paderewski is a musician. Imagine further that Peter believes that Paderewski the musician has red hair. Picture a further individual, Petra, who is also confused about Paderewski. She believes the same uncoordinated token propositions as Peter, but she believes that it is Paderewski the politician who has red hair. This would generate yet further token propositions, since the propositions Petra believes do not stand in the same coordination relations as those that Peter believes. We could continue this forever: Imagine individuals who believe the same set of Millian/Russellian propositions, but that differ in their coordination relations. People make mistakes like this all the time, and for every Millian/Russellian proposition there must be enough token propositions to account for these kinds of mistakes. Thus, by appealing to tertiary content we are committed to the existence of possibly indefinitely many token propositions *in addition* to classical Millian/Russellian propositions.

We are beginning to see that the Semantic Relationist framework is extremely complex. In order to give a full account of cognitive significance, the Semantic Relationist is forced to posit a very complex semantics and with that comes a complex metaphysics: There are three levels of content and the third level



generates indefinitely many abstract entities, namely token propositions. Fine's framework is supposed to provide a way to explain coordination within a general Millian/Russellian framework without having to appeal to the notion of Fregean senses or the like. However, one of the most attractive features of Millian/Russellianism is that it comes with minimal propositional complexity and commitments. Even though Fine's framework does not invoke the notion of descriptive senses, it invokes the notion of 'guises' and these (whatever they are) exist at a different semantic level than classical Millian/Russellian propositions. As a result, I suspect that Fine's view would not be compelling to most Millian/Russellians, including those who are sympathetic to the core commitments of relationism. Moreover, if we have an alternative framework that is (at least) equally explanatorily powerful but that does not need to postulate such a complex semantics, this framework would be preferable.

I think Vehicle Relationism is such a framework. As long as we take thinking to be a triadic relation it is possible that a thinker may stand in relation to the same proposition in two different and numerically distinct ways. A thinker may believe the very same proposition twice but in different ways. All this amounts to is just for the thinker to have two sequences of mental tags that are not connected by pointers but that happen to express the same proposition. Similarly, we may allow that one thought may negate what the other states, as long as we have the option of appealing to a difference at the level of representational vehicles. Importantly, Vehicle Relationists do so without having to postulate multiple levels of semantics or duplicating propositions.

I've already illustrated the explanatory power of the Vehicle Relationist framework when it comes to solving some of the central puzzles within the philosophy of mind. If both frameworks are equally suited to explain such puzzles we ought to choose the framework with fewer metaphysical commitments and this, I have argued, is Vehicle Relationism. In what follows I will provide a further argument in favour of Vehicle Relationism. I will question the cognitive effect of semantic coordination and argue that the Semantic Relationist must posit something akin to pointer relations in order to account for how coordination relations can be cognitively causally efficacious. If the

Semantic Relationist needs pointers and the Vehicle Relationist does not need semantic relations, Vehicle Relationism is the better option since the framework is more parsimonious.

### *5.3. Semantic Relationism: Cognitive Effect*

The reason why classical Millian/Russellians face problems with Frege cases is that sameness or difference in Millian/Russellian propositions is non-transparent, and as a result individuals may be mistaken about such sameness or difference in propositional content. That is, as we have seen a person can believe the same Millian/Russellian proposition twice without being aware of this. Put differently, two thoughts may have the same Millian/Russellian propositional content but nonetheless play different roles in the cognitive life of a single individual. Coordination relations, however, are not the kind of things thinkers can be mistaken about in this way. That is, necessarily if two thoughts are positively coordinated a thinker treats them as concerning the same referent.<sup>21</sup>

According to the Semantic Relationist, coordination relations are abstract objects of a certain sort. They are constituent parts of propositional contents. As such, they are not really “in the mind” of the thinker. In contrast, the Vehicle Relationist takes such relations to be elements of mental representations; they are proper constituents of the language of thought. As such, they have physical-causal (presumably neurological) properties in addition to semantic properties (c.f. Fodor 1987).

In order for semantic relations to account for coordination, it must be possible for such relations to have a causal effect on the cognitive workings of individuals.

---

<sup>21</sup> Importantly, the claim here is not that one has infallible higher-order knowledge regarding the coordination relations of first order-beliefs. That is, it may be possible for someone to have positively coordinated first-order beliefs, and then at a later time wonder whether or not those beliefs actually concerned the same referent without violating rational norms. Rather, the claim is that, if two thoughts are positively coordinated, the cognitive system will treat them as such. Likewise, if two thoughts are negatively coordinated, the cognitive system will treat them accordingly. In other words, one cannot have positively coordinated beliefs without also being rationally disposed to trading on identity.

If they didn't, we could imagine someone believing a positively coordinated proposition and fail to have the sort of behavioural and rational dispositions that are associated with positive coordination, such as a disposition to trade upon identity. In other words, thinkers cannot wrongly have a positively (or negatively) coordinated thought. If they could, we would have failed to account for Frege's puzzle, since someone could have two thoughts whose content is the same coordinated proposition but that nonetheless play different roles in cognition. We would then have to introduce a further notion in order to explain the difference in cognitive significance of two thoughts that express the same coordinated proposition.

The question, then, is this: what warrants the assumptions that semantic relations are of the right kind to have the relevant causal effects in cognition? Fine would have to say that sameness or difference of uncoordinated (i.e. primary) propositions is not manifest, but that somehow sameness or difference of coordinated (i.e. secondary and tertiary) propositions is manifest to thinkers. This would, however, be very puzzling. Propositions – even the coordinated ones – are abstract and mind-external on Fine's view. As such, they seem to fall outside of the realm of physical causation. How, then, can our standing in relation to such entities help explain our cognitive dispositions? We all agree that a thinker may fail to know that two uncoordinated (i.e. primary) propositions co-refer. But what is it about the nature of coordinated (i.e. secondary and tertiary) propositions that makes it possible for us to have infallible cognitive access to them when we do not have such access to uncoordinated propositions, given that both sorts of propositions are abstract entities?

Fodor makes a similar point regarding Fregean senses. He considers the claim that senses, which are mind-external, abstract objects, are such that they can have causal effects on cognition:

The question thus arises what, if anything, is supposed to legitimize this assumption. As far as I can tell, unless you're prepared to give up

[the thesis that senses are abstract and hence non-mental], the only answer a Fregean theory allows you is: sheer stipulation. (1998, 17)

The point holds for all theories that take coordination to be explained in terms of non-mental, abstract entities. We have no reason to believe that such entities are of the right kind to have the sort of causal effect on cognition needed to explain coordination. In particular, anyone who accepts that sameness or difference in classical Millian/Russellian propositions is not manifest to thinkers, as Fine does, would have to provide reasons to think that sameness or difference of other abstract entities, such as for instance Fine's token propositions, can be manifest to thinkers. As for now we have no reason to think that this is the case.

The only plausible explanation of how sameness or difference in token propositions (i.e. Millian/Russellian proposition plus coordination relations) can be manifest to thinkers, as far as I can see, is that the coordination relations must somehow be reflected in something that is of such a nature that they may enter into cognitive causal relations, namely mental representations. Again, Fodor makes the same point:

Whatever distinguishes coextensive concepts is ipso facto 'in the head'. This means, something like that it's available to be a proximal cause (/effect) of mental processes. (Fodor 1998, 15)

This, of course, does not rule out the possibility of there being a difference of coextensive concepts both in representational vehicles and in the semantic properties of concepts. What's important is that at the very least, in order to account for the possibility of co-referential concepts playing distinct roles in cognition, there must be a difference in mental representations.<sup>22</sup> As a result, it seems that in order to give an account of the cognitive role of concepts and thoughts, the semantic relationist also needs to appeal to coordination of mental representations.

---

<sup>22</sup> Laurence & Margolis (2007) use a similar argument in order to argue that we need mental representations in addition to propositional content.

In fact, Fine says that he does not “wish to deny that the semantic relationship – of representing-as-the-same – might hold in virtue of a syntactic relationship” (Fine 2007, 41).<sup>23</sup> He says this in connection to coordination in language, but later on he draws a close connection between the representational character of language and thought:

For one thing, it is not at all clear to what extent there *are* two systems of representation. [...] even if the systems of representation are by and large disjoint [...] how can the *vehicle* of representation – be it speech or writing or thought – make any difference to its representational character? (Fine 2007, 77)

Hence, it seems plausible that if coordination in language holds in virtue of syntactic properties, so does coordination in thought.

This, in and of itself, does not commit the Semantic Relationist to Vehicle Relationism specifically – there are other alternatives to how we may explain coordination of mental representations, such as Fodor style LOTH. However, Fine himself highlights the non-transitive nature of coordination relations when explaining the intersubjective puzzle stemming from the puzzle of Paderewski (Fine 2007, 105-115). For this reason, it seems that Vehicle Relationism would be a more suitable framework.<sup>24</sup>

If this is correct, it seems that the Semantic Relationists must posit something akin to pointer relations in order to give a full account of coordination in thought: The fact that coordination relations must be such that they can enter into causal relations in cognition implies that coordination relations must be instantiated at the level of mental representations. If coordination relations are non-transitive, the pointer picture is the best framework for giving an account of how such coordination relations are implemented.

We see that Vehicle Relationism and Semantic Relationism are not mutually exclusive. However, it seems that postulating a complex relationist semantics

---

<sup>23</sup> Again, “representing-as-the-same” just means that there is positive coordination.

<sup>24</sup> See section 4.1. above.

would be unmotivated. If we accept Vehicle Relationism, and Vehicle Relationism does the relevant explanatory work, semantic relations do not add anything of explanatory value when it comes to the explanation of coordination in thought. Hence, I suggest that our best option is to reject Semantic Relationism about coordination in thought in favour of Vehicle Relationism plus (classical) Millian/Russellianism about propositional content.

In the next and penultimate section of this paper I will set out some objections to what I've been arguing this far. I will suggest some replies to these worries.

## 6. Objections and Replies

Vehicle Relationism is a novel account of coordination relations. I've spelt out the central claims of the view and compared it to other related and competing frameworks. Of course, more work is needed before we can be certain that the framework provides a satisfactory account of coordination relations in thought. In what follows I will consider some objections and respond to them in turn. I will signal where more work is needed.

*Objection:* You have not said enough about the nature of the pointer relations. What are they and in virtue of what do they obtain? The general notion seems highly mysterious.

*Reply:* The notion of pointers is a functional one. The cognitive system, I claim, essentially functions so as to store information based on what is taken to be sameness or difference in reference (c.f. Campbell 1987, see also section 2.3. above). There are all kinds of pragmatic explanations of why co-reference is recognized in certain cases and not in others. The main claim is that an essential part of information processing involves categorizing and recognizing sameness or difference between the object of a new input belief and the objects of one's already existing propositional attitudes. In this respect, the framework is similar to Fodor's language of thought hypothesis, according to which information is labelled according to sameness or difference in reference and to the mental file

framework (discussed in chapter 1), according to which information is sorted into files based on recognition of sameness or difference in reference.

*Objection:* The pointer picture, as you admit yourself, is in many ways similar to other more well-known frameworks, such as Fodor's LOT and the mental file framework. It seems more convenient to use the notion of sameness of syntactic form or sameness of files to talk about these relations, so why should we use the more complex language that comes with the pointer metaphor?

*Reply:* I discussed one reason to not use the classical Fodorian LOT earlier in this paper, when I presented an argument to the effect that coordination is non-transitive. This also provides a reason to reject the mental file framework. Further, in Chapter 1 I argued that the file framework fails to account for coordination in the case of relational predicates. Hence, I think that strictly speaking none of these frameworks can provide a satisfactory and unified account of coordination in thought. That said, the notion of a syntactic form or that of a mental file do come in handy in many cases. In normal cases of coordination (i.e. cases not displaying non-transitivity) of one-place predicates it is often helpful to think of cognitive processing in terms of files or the like. However, my point is that, strictly speaking, such metaphors only take us so far: when things become more complex we need the more general notion of pointer relations. Importantly, the simple cases can also be explained in terms of pointer relations, and so Vehicle Relationism provides a general unified account of coordination.

*Objection:* You are not really solving the problem of coordination in thought. Rather, you are merely re-describing the phenomenon.

*Reply:* I have provided an account of how it is that coordination is realized in thought. Coordination relations, recall, are just those relations that obtain whenever someone is rationally warranted in trading on identity. Note that I am not claiming that these relations are primitive. Rather, I argue that coordination relations hold in virtue of further relations – relations that are understood as proper parts of the language of thought. These relations, I have argued, are

representationally primitive,<sup>25</sup> but that is not to say that coordination relations themselves are primitive.

This account differs from the traditional accounts of coordination that explain such relations in terms of intrinsic representational features of concepts and thoughts. The difference, however, lies in what the frameworks have to say about in virtue of what such coordination relations obtain, rather than whether or not we can give an account of such coordination relations. I have shown that Vehicle Relationism solves some of the central puzzles within the philosophy of mind. Given the explanatory power of the framework, it would be a mistake to insist that coordination in thought must, somehow, ultimately be reducible to intrinsic representational features of concepts and thoughts.

*Objection:* You only talk about thoughts, whereas some of the competing views, such as Semantic Relationism, also provide an account of coordination in language. Without further evidence that your framework can enlighten us on coordination in language, what reasons do we have to prefer your framework to that of e.g. Semantic Relationism? You might need semantic relations in the case of language, so maybe the two versions of relationism are not that different in complexity after all.

*Reply:* It is beyond the scope of this paper to give an account of coordination in language. Vehicle Relationism is in theory compatible with any account of coordination in language. Even so, some things seem to follow naturally (although not necessarily) about the semantics of natural language from what I say about the semantic content of thoughts. When it comes to thoughts, I would say that two thoughts such as NO ONE WHO KNOWS THAT GREEKS ARE GREEKS DOUBTS THAT GREEKS ARE GREEKS and NO ONE WHO KNOWS THAT GREEKS ARE GREEKS DOUBTS THAT GREEKS ARE HELLENES<sup>26</sup> have the same semantic content and so they have the same truth conditions. Even so, they may play different roles in cognition due to a difference in pointer relations. When it comes to

---

<sup>25</sup> That is, they cannot be reduced to sameness or difference in intrinsic representational features of concepts and thoughts. The pointers are a representationally irreducible part of the language of thought.

<sup>26</sup> Cases such as these are generally known as *Mates cases* (see Mates 1952).



sentences expressing such beliefs, I would therefore find it natural to say the same thing: The sentences have the same truth conditions. Either both are true or both are false. In the actual world, both sentences are false, since there are clearly people who doubt that Greeks are Hellenes. Even so, the sentences clearly differ in syntax, since they contain words of different types. This difference is enough to explain why someone might come to think that one is true while the other is false, just like someone who hears the two sentences “Hesperus is a planet” and “Phosphorus is a planet” might come to think that one is true while the other is false. However, our epistemic dispositions need not reflect an actual difference in semantic content.

I am, of course, aware that this view of linguistic meaning is controversial; much of the discussion within the philosophy of language since Frege, even within the Millian/Russellian tradition, has focused on exactly how to avoid this conclusion. I find the parsimonious ontology of the view appealing, but Vehicle Relationism is by no means committed to this specific view of linguistic meaning. The question, however, still stands: By adopting this view of linguistic meaning, is Vehicle Relationism as explanatorily powerful as Semantic Relationism when it comes to coordination in language? Only future development of the Vehicle Relationist framework will show. Perhaps all we need for coordination in language is formal relations à la Heck (2012). For what it’s worth, I suspect this might be the case.

## 7. Conclusion

I have proposed a novel framework for understanding the nature of coordination in thought. According to this framework coordination in thought is to be explained in terms of representationally primitive relations that have representational vehicles (i.e. mental tags) as relata. On this framework, we get a natural distinction between manifest and opaque co-reference as well as between manifest and apparent co-reference. Such distinctions are important for the explanation of cognitive significance and psychological generalizations.

I have shown how the Vehicle Relationist framework can solve some of the long-standing puzzles within the philosophy of mind pertaining to cognitive significance. The explanation in terms of pointer relations is compatible with classical Millian/Russellianism about propositional content.

I have also compared the suggested framework to similar views. The proposed framework draws on insights from relationists such as Heck and Fine as well as Language of Thought theorists such as Fodor, but combines these elements in a novel way. I have also argued that Vehicle Relationism is preferable to these competing views of coordination in thought.

In general, the framework of Vehicle Relationism provides a highly parsimonious account of coordination in thought. At the same time, the framework is explanatorily powerful. This gives us reason to think that Vehicle Relationism is the best contender for an account of coordination in thought.



## Originalism and Coordination in Thought: In Defence of Vehicle Relationism

*According to the Publicity Constraint on concept individuation, the nature of concepts must be such that distinct individuals may use the same concept or concept tokens of the same type. At the same time, concepts are taken to explain intrapersonal cognitive capacities, such as rational reasoning. This indicates that a difference in cognitive role must be reflected in a difference in concepts or concept types. This is the Fregean Constraint on concept individuation. There is a tension between the Publicity Constraint and the Fregean Constraint. The former seems to require a coarse-grained individuation of concepts while the latter requires fine-grained individuation conditions. In this paper I assess a recent account of the individuation of concepts that purports to overcome such difficulties. According to this view, Originalism, concepts are public and sharable but at the same time, the framework is specifically developed to account for intrapersonal rational reasoning. I argue that the theory fails to account for the cognitive role of concepts. I then suggest a minimal addition to the framework that allows it – or more generally any theory that takes concepts to be public – to account for an individual's rational cognitive capacities.*

## 1. Introduction

It seems intuitively plausible that distinct individuals can and must have thoughts that in one way or another are *the same*. For instance, if you and I both believe that Bob Dylan is American it seems natural to say that we have the same belief. Many philosophers have traditionally taken the job description for thoughts and concepts to require the nature of concepts to be such that distinct individuals may entertain the very same concept or concept tokens of the same type. One task assigned to concepts that seems to require them to be public and shareable is their contribution to the explanation of communication.<sup>1</sup> Heck (2002) puts what they call the naïve conception of communication thus:<sup>2</sup>

What is the purpose of communication? At a minimum, it would seem, part of its purpose is to transfer information from one speaker to another: I have a belief, or take myself to know something, and I want to get you to believe the same thing. [...] when you grasp the content of my assertion, you thereby grasp the very Thought I believe and am trying to communicate to you. [...] when I communicate, I am trying to bring it about that someone else should come (to have the opportunity) to share a belief with me [...]. (Heck 2002, 6)

If this is correct, communication requires thoughts and concepts to be public, i.e. their nature must be such that distinct individuals can share concepts and thoughts. We thus get the following constraint on concepts:

---

<sup>1</sup> Other such tasks include accounting for psychological generalizations (e.g. Fodor 1998, 28—29, Prinz 2002, 14—15) and accounting for agreement and disagreement (e.g. Sainsbury & Tye 2012, 21).

<sup>2</sup> Heck says that “there is evidence that Frege held such a view of communication” (Heck 2002, 7) and points to Frege’s (1892, 159—162) arguments against the claim that thoughts are ideas. Without going into Frege exegesis, this seems correct. Frege says that senses ‘[...] may well be common property of many and is therefore not a part or mode of the single person’s mind: for it cannot well be denied that mankind possesses a common treasure of thoughts which is transmitted from generation to generation.’ (Frege 1892, 188).

**Publicity Constraint:** The nature of concepts is such that they can be shared between distinct individuals.<sup>3</sup>

The publicity constraint is accepted by many philosophers (e.g. Peacocke 1992, Rey 1994, Fodor 1998, Sainsbury & Tye 2012). This constraint, however, seems to be in tension with another plausible assumption about the nature of concepts and thoughts.

Concepts play an important role when it comes to explaining our cognitive rational workings. Concepts are the constituents of thoughts, and the inferential relations between thoughts are closely related to features of their constituent concepts. If two thoughts are referentially equivalent, but nonetheless play different roles in cognition, this is usually taken to be due to a difference in the constituent concepts of the two thoughts. Hence, concepts play an essential explanatory role when it comes to rational intrapersonal processes. The nature of concepts must thus be such that it reflects individuals' cognitive workings.

It is commonly agreed that it is possible for an individual to have two singular thoughts concerning the same referent without the sameness of reference being transparent to the thinker. Frege's (1892) puzzle of Hesperus and Phosphorus illustrates this: the Ancient Babylonians used the term 'Hesperus' to refer to the brightest visible object on the evening sky and the term 'Phosphorus' to refer to the brightest object visible on the morning sky. Unbeknownst to them, 'Hesperus' and 'Phosphorus' refer to the same heavenly body. The Ancient Babylonians may have had beliefs about Hesperus without having corresponding beliefs – or perhaps even contradictory beliefs – about Phosphorus. They were warranted in this, since the co-reference of the concepts HESPERUS and PHOSPHORUS were not manifest to them. Frege's observation puts the following constraint on the nature of concepts and thoughts:

---

<sup>3</sup> c.f. Onofri 2016, Duhau 2012.

**The Fregean Constraint:** If a subject can rationally believe of  $x$  that it is  $F$  under thought  $t_a$ , and at the same time believe of  $x$  that it is not  $F$  under  $t_b$  then  $t_a$  and  $t_b$  are different thoughts.<sup>4</sup>

How best to account for the lack of manifest co-reference has been a central question in the philosophy of mind and language since Frege. The pairs [HESPERUS, HESPERUS] and [HESPERUS, PHOSPHORUS] contain concept tokens with the same referent, but only in the former pair is the co-reference manifest. Someone who has two beliefs, each of which contains a token of HESPERUS, may exploit this manifestness of co-reference and combine the two beliefs in a generalization. For instance, if someone believes both that Hesperus is visible and that Hesperus is bright, they may infer that something that is bright is visible. They may do so directly, without having to make any further judgements to the effect that Hesperus is Hesperus. This is what Campbell (1987) calls *trading on identity of co-reference*. I will say that two concept tokens that are manifestly co-referential in this way are *positively coordinated*.

In contrast, two concept tokens that are not manifestly co-referential are *negatively coordinated*. In cases where co-reference is not manifest, the thinker is not rationally warranted in trading on identity when making generalizations without making a further identity judgement. For instance, if someone has the belief that Hesperus is visible and also believes that Phosphorus is bright, she is not warranted in inferring from these two beliefs alone that something that is bright is visible. In order to make this generalization, she would also have to believe that Hesperus is Phosphorus. Since someone may rationally believe that Hesperus is visible and at the same time doubt that Phosphorus is visible, it follows from Frege's constraint that the concepts HESPERUS and PHOSPHORUS must be distinct concepts, since the two thoughts are otherwise structurally identical.

Like the Publicity Constraint, then, the Fregean Constraint puts restrictions on the individuation of concepts: it requires concepts to be such that they can account for intrapersonal coordination relations. However, as several authors

---

<sup>4</sup> c.f. Onofri 2018, 6.

have argued (e.g. Crimmins 1992, Heck 2002, Laurence & Margolis 2007, Duahu 2012, Onofri 2016) there is a tension between the Fregean Constraint and the Publicity Constraint. In order for concepts to be public and shareable their individuation conditions must be coarse-grained so that it is possible for distinct individuals to have the same thought or thoughts of the same type. For instance, it seems that you and I may both have the belief that Bob Dylan is American even though we have very different conceptions and beliefs regarding Dylan himself. Hence, in order for concepts to be public, their nature must be such that sameness or difference of concepts does not fully depend on the mental lives of single individuals.

At the same time, it seems that in order for concepts to account for intrapersonal coordination relations their individuation condition must be fine-grained. That is, the nature of concepts must be such that it reflects how rational subjects reason. Consider again the case of Bob Dylan. Let's say I am confused and come to think that there are two distinct people named 'Bob Dylan'. I come to think that Dylan the musician is a different person than Dylan the Nobel Price winner. I may then rationally believe the proposition that Dylan won a Nobel Price and at the same time doubt that Dylan won a Nobel Price. But, then, from the Fregean Constraint it follows that the two concepts  $DYLAN_{(the\ musician)}$  and  $DYLAN_{(the\ Nobel\ Price\ winner)}$  are distinct concepts. At the same time, from the publicity constraint it seems that each of my DYLAN concepts are the same as your DYLAN concept, since I may successfully communicate with you using either concept.

Cases like these are well-known and were first introduced by Kripke (1979). They create a problem for those who want to claim both that concepts are public and at the same time that they explain intrapersonal coordination. Hence, it seems that we are forced to abandon one of the constraints. According to a recent theory of concepts, however, such complications can be overcome. Sainsbury and Tye (2011, 2012) propose a novel account of the nature of concepts, which supposedly does justice to both constraints. According to this view, *Originalism*, concepts are public and so it follows naturally that distinct individuals can share concepts. At the same time, the view is specifically developed to solve some of the longstanding puzzles within the philosophy of



mind pertaining to intrapersonal cognitive capacities of individuals. If the view is successful in doing this, it would be exceedingly promising, since it would do justice to both the Publicity Constraint and the Fregean Constraint.

In this paper, however, I will argue that Originalism fails to give a satisfactory account of the cognitive role of concepts. According to Originalism concepts are individuated by their historical origins. On this view, the cognitive role of concepts is explained in terms of sameness or difference in concept types. At the same time, however, traditional puzzles such as Kripke's (1979) Paderewski case force the originalist to say that such sameness or difference in type is not transparent to thinkers.

I will point to a structural problem with this kind of view. I will argue that Originalism – or more generally any theory that takes the cognitive role of concepts to be accounted for in terms of types of concepts, but at the same time holds that sameness or difference in concept types is not transparent to thinkers – fails to account for rational reasoning.

I will suggest a minimal addition to such frameworks that allows them to account for a thinker's rational cognitive capacities. I propose that the cognitive role of concepts is to be accounted for in terms of primitive relations that obtain between representational vehicles rather than in terms of features such as representations being of the same type. This opens up the possibility of having coarse-grained individuation conditions for concepts (and thus respect the Publicity Constraint on concept individuation), and at the same time do justice to Frege's observations without strictly speaking conforming to the Fregean Constraint.

In section 2, I present the key claims of Originalism. In section 3, I assess the Originalist account of the cognitive role of concepts by looking at how Originalists explain the puzzle of Hesperus and Phosphorus as well as the puzzle of Paderewski. In section 4, I argue that the originalist account of the cognitive role of concepts is untenable. In section 5 I make a positive proposal as to how originalists – or, more generally, any theorist who takes concepts to be public –

may account for rational reasoning by making a minimal addition to their theory. In short, the suggested view takes sameness or difference of concept types (individuated coarsely) to be non-transparent to thinkers while explaining rational reasoning in terms of primitive relational features of mental representations rather than in terms of sameness or difference in concept types. Finally, I consider a similar move suggested in the literature (Prosser 2018) according to which the explanation of coordination in intrapersonal cases may come apart from coordination in interpersonal cases. This view differs from the proposal made in this paper in that it takes the Fregean Constraint to be accounted for in terms of sameness or difference of concepts, whereas communication is explained in terms of relational aspects of thoughts. In contrast, according to the framework proposed in this paper, the Publicity Constraint is accounted for in terms of sameness or difference of concepts, whereas the Fregean observations are accounted for in terms of relational features of concepts and thoughts.

## 2. Originalism

Originalism is specifically branded as providing solutions to long-standing puzzles within the philosophy of mind pertaining to intrapersonal coordination. At the same time concepts are taken to be shareable in that distinct individuals may use concepts of the same type. In what follows I will present the key claims of Originalism. I start out by making clear what exactly the Originalists understand by ‘concepts’.

Originalists take concepts to be

mental representations of a sort deployed in thought; they are representational constituents of thoughts. Thoughts are made up of concepts, and what thoughts as a whole represent is a function of their component concepts: what they represent and how they are combined. [...] Since concepts represent things and they make up thoughts, they too have representational contents. (Sainsbury & Tye

2012, 1)

Concepts are vehicles of representation, tools for thinking (Sainsbury and Tye 2011, 101).

On this view, thinkers do not stand in direct relation to the content of the concepts deployed in thought. Thinkers are directly related to concepts (understood as vehicles of representation) and related to the content of such concepts via such representational vehicles. Originalists hold that “cognitive processing depends not directly on content but on the vehicles of content: concepts and thoughts” (Sainsbury and Tye 2012, 57). This opens up the possibility of two concepts having the same content, but nonetheless playing different roles in cognition due to a difference in the representational vehicles.

Since concepts are the constituents of thoughts, the same point applies to thoughts as well: “distinct thoughts, even if they are referentially isomorphic, can play different cognitive roles” (Sainsbury and Tye 2011, 101—102). This is how originalists account for the Fregean constraint. I return to this in part 3 below, where I spell out the originalist solution to some of the central puzzles within the philosophy of mind.

When an individual deploys a given concept in thought, she entertains a concept *token*. Concept tokens enter into causal computational relations in cognition. In contrast, concept *types* are non-eternal abstract objects:

Thoughts are structures of concepts, so if one can say what thoughts are one is well on the way to saying what concepts are. Thoughts are abstract things that can be evaluated as true or false, can be believed or doubted, stand in logical relations, can be shared by different thinkers and can typically be expressed by indicative sentences. [...] Non-eternal abstract continuants [...] is the category to which concepts belong. (Sainsbury & Tye 2012, 63)

Since, on this view, concepts and thoughts are abstract objects it is possible for distinct individuals to use (i.e. token) concepts and thoughts of the same type.

Although different individuals use distinct thought tokens, the publicity constraint can be respected by an appeal to sameness or difference in types. The concept types are thus public and as such “concepts are typically shareable” (Sainsbury and Tye 2012, 59).<sup>5</sup> Individuals have their concepts in virtue of being part of a language community, and the participants in the language community share concepts.<sup>6</sup>

We now have some understanding of the originalist account of the metaphysics of concepts and thoughts. A further question concerns the exact nature of concept types. More specifically, what are the individuation conditions for concepts?

As mentioned, originalists hold that concepts (types) are non-eternal entities. They come into existence at a particular time in history and they may go out of existence at a later time. The point in history at which a given concept comes into existence is the originating use of that concept. The key claim of originalism is that two concept tokens are of the same type if and only if they have the same historical origin:

**Originalism:** Concept C1 = concept C2 if and only if the originating use of C1 = the originating use of C2 (c.f. Sainsbury & Tye 2011, 105).

On this view, it is the case that for every concept there is just one originating use and that every originating use of a concept is the origin of one concept only (c.f.

---

<sup>5</sup> The exception is indexical concepts, which are not shared between distinct individuals. Even so, such indexical concepts are shaped by shared requirements: “Indexical concepts are copied from templates that impose more or less specific requirements on the use of the concept they shape” (Sainsbury & Tye 2012, 85). Importantly, “a concept-template is not a concept, but rather a recipe for forming concepts, a pigeon-hole to hold many different but similar concepts” (Sainsbury & Tye 2012, 51). Hence, indexical concepts are not shared between distinct individuals, but they are similar due to being shaped by the same requirements.

<sup>6</sup> Note that young children may come to form their own individual concepts when interacting with the world. For instance, they may form a specific concept when interacting with cats. However, the individual concepts children acquire at a young age will typically be replaced by public concepts when the children interact with others in their language community. Sainsbury and Tye take children’s willingness to accept correction to be an indication that children replace their individual concepts with public concepts at some point in early development (2012, 60). On the originalist account, the concept a child has for picking out cats, introduced independently of other participants in her language community, is distinct from the concept CAT she uses after having acquired the public concept.

Sainsbury & Tye 2011, 104).<sup>7</sup>

There are two sufficient conditions for a use of a concept to be non-originating:

- 1) The use involves deference to other uses, by the same subject or other subjects.
- 2) The use involves informational accumulation from other uses, by the same subject or other subjects. (c.f. Sainsbury & Tye 2011, 102)

If a concept token belongs to a chain of deference and is not itself the starting point of such a chain, the use is a non-originating use. That is, if an individual intends to use a concept the same way as others in her language community, her use is a non-originating use.

On this view, knowledge of the content of concepts is not necessary for someone using a given concept: “Concept possession is consistent with all sorts of mistakes and misunderstandings about the concept’s subject matter” (Sainsbury & Tye 2012, 55). Further, Originalists “have no room for a notion of the “correct” use of a concept [...] for originalism there is simply the question whether a subject uses or does not use a concept on an occasion. If it is used at all, then it is used “correctly”” (Ibid., 85). If the subject defers to other uses of a given concept this guarantees that she uses that concept, even if she is mistaken about the subject matter. Hence, on this view, concept individuation does not depend on details about the inner workings of single individuals. The result is that concepts are individuated in a coarse-grained way, making it possible for distinct individuals to have concept tokens of the same type. Originalism thus adheres to the Publicity Constraint on concept individuation.

---

<sup>7</sup> A complication for the Originalist theory is that some concepts often are taken to have more than one originating use. In order to account for this, the Originalists introduce the notion of *conceptual fusion*. In the case of conceptual fusion, two (or more) concepts fuse into one concept. At the time of a conceptual fusion the concepts that fuse together go out of existence and a new concept comes into being. The new concept originates at the point of fusion. Another complication for Originalism is that two or more concepts often are taken to have the same origin. In order to account for this, the Originalists introduce the notion of *conceptual fission*. In the case of a conceptual fission, one concept fissions into two (or more) concepts. In this case, the original concept is of a different type than the new concepts that come into being. The new concepts that come out of the fission have their origin in the first intentional use of each of the concepts introduced by the fission (c.f. Sainsbury & Tye 2012, 66—68).

To sum up, originalists take concept tokens to be representational vehicles, constituents of thoughts. Such concept tokens enter into causal relations in cognitive computation. Concept types are abstract entities that are publicly available in such a way that distinct individuals may use concepts of the same type. Concepts are individuated by their historical origins. In this way, concepts are similar to words in language; words are individuated by their etymological history and different speakers may use words of the same type. This is how the Originalists account for the Publicity Constraint. In the next section, I will take a closer look at how the Originalists do justice to the Fregean Constraint by looking at their proposed solutions to puzzles pertaining to the cognitive role of concepts and thoughts.

### 3. Originalism and the Fregean Constraint

In this section we'll see in more detail how originalists account for the Fregean Constraint. There are two puzzles in particular that are of special interest when it comes to the Fregean Constraint. First, there's Frege's (1892) own puzzle of Hesperus and Phosphorus. Second, there's Kripke's (1979) puzzle of Paderewski. I'll present the originalists' proposed solution to each puzzle in turn.

#### *3.1. The Puzzle of Hesperus and Phosphorus*

Frege's (1892) puzzle of Hesperus and Phosphorus gave the foundation for the Fregean Constraint. According to the Fregean Constraint, recall, a rational individual can only think of the same referent that it is F and that it is not-F at the same time if the individual thinks of the referent in different ways, i.e. by deploying distinct thoughts. This observation coincides with what Sainsbury & Tye call a *Fregean Datum*:

A Fregean datum is that it's one thing to think that Hesperus is Hesperus, and another to think that Hesperus is Phosphorus; one

thing to think that Hesperus is visible, another to think that Phosphorus is visible. (Sainsbury & Tye 2012, 53)

Originalists accept this, and it follows that the beliefs HESPERUS IS VISIBLE and PHOSPHORUS IS VISIBLE must be distinct thoughts, since rational individuals may form one of the beliefs while rejecting the other:

We agree [with the Fregean datum]. Different thoughts are involved, that is, different structures of concepts, since the concept HESPERUS is distinct from the concept PHOSPHORUS (Ibid.).

Even though the two concepts refer to the same object, they may play different roles in cognition due to them being of different types. The originalists disagree, however, with Frege's (1982) conclusion that

the difference requires postulating any additional semantic layer. [...] Distinct concepts can, and typically will, play different roles in our cognitive activities, even if they have the same content. [...] The work supposedly done by difference of sense can be done better by difference of concepts. (Sainsbury & Tye 2012, 53—54)

Hence, on the Originalist framework we may account for the Fregean data in terms of a difference in types of representational vehicles rather than a difference in semantic content.<sup>8</sup>

According to the Originalists, the concepts HESPERUS and PHOSPHORUS are of different types as a result of them having distinct historical origins. The Ancient Babylonians introduced HESPERUS in the evening, while PHOSPHORUS was first used in the morning. Hence, the originating use of HESPERUS ≠ the originating use of PHOSPHORUS. Since referentially isomorphic thoughts that differ only by

---

<sup>8</sup> Fodor (2008) agrees with the originalists that Frege cases can be explained without appeal to semantics that goes beyond reference. However, the two theories disagree on several matters. One such matter is this: While originalists take concepts to be individuated historically, Fodor thinks that such entities are of different types "when they differ in the (presumably physical) properties to which mental processes are sensitive" (Fodor 2008, 79). On Fodor's view, subjects cannot be mistaken about the type and number of concepts deployed in thought. In (3.2) we'll see that originalists disagree; they deny that concepts are transparent to the thinker. In part 4 I argue that this claim creates problems for the Originalist account of coordination in thought.

the substitution of HESPERUS for PHOSPHORUS are of different types, the two thoughts, if tokened by a single individual, may play different roles in cognition. This is why someone may rationally endorse one of the thoughts while at the same time rejecting the other.<sup>9</sup>

In the case of Hesperus and Phosphorus, then, the Originalist account of concepts successfully respects the Fregean Constraint on concept individuation since it seems plausible that the concepts in question (i.e. HESPERUS and PHOSPHORUS) have distinct historical origins. This, however, may just be a historical coincidence. If sameness or difference in concept types are to account for the Fregean data the individuation conditions of concepts need to be such that all concept tokens that play distinct roles in cognition come out as distinct types. This means that if two concept tokens play distinct roles in the mind of a single individual they better be tokens of concept types with distinct origins. In the specific case of HESPERUS and PHOSPHORUS the originalist is lucky, because it seems plausible that the concepts were introduced at different points in history. However, there are other puzzles, such as Kripke's (1979) puzzle of Paderewski, which renders the appeal to distinctness of concept types impossible for the Originalist. This forces the Originalist to pursue a different strategy. In the next section I will present the Originalists' proposed solution to this puzzle.

### *3.2. The Puzzle of Paderewski*

Kripke's (1979) puzzle of Paderewski runs as follows: Ignace Paderewski was a popular Polish pianist. He was also engaged in politics and after the First World War he became the Polish prime minister. Now consider a person, Peter, who independently comes to know about Paderewski the piano player and Paderewski the politician without realizing that they are in fact the same person.

---

<sup>9</sup> Originalism coincides with Ruth Millikan's theory of concepts in certain respects: Millikan agrees with Sainsbury and Tye that Fregean data are to be explained by appeal to sameness and difference in vehicles of content rather than the content expressed by such entities. They also agree that concepts are to be individuated by way of their historical properties. However, while originalists take concepts to be public, Millikan thinks concepts are individual and not sharable: "I have concepts and you have completely other concepts, though many of them may be concepts be of the same thing" (Millikan 2011, 6).



Instead Peter believes that it is a case of two different people sharing the same name. Having been to one of Paderewski's concerts, Peter comes to believe that Paderewski has musical talent. However, Peter also believes, on good authority, that no politicians have musical talent and that no pianists are politicians. Thus it also seems plausible to attribute to Peter the belief that Paderewski lacks musical talent. The puzzle, then, is this: Peter is rational, but he seems to have contradictory beliefs; how can this be?

The problem is that, if we assume that concepts are public, there should be only one concept PADEREWSKI. Hence, when Peter uses the concept PADEREWSKI at the concert he uses the very same concept as others in his language community. At the same time, when Peter uses the concept PADEREWSKI at a rally, Peter is also using the same concept as others in his language community. Hence, by transitivity, Peter is using the same concept both when forming his belief that Paderewski has musical talent and also when forming the belief that Paderewski does not have musical talent. But then Peter's beliefs are contradictory.

When Peter forms his beliefs he defers to other uses in his language community. According to the Originalist, this ensures that he uses the same concept as the others. There is only one public concept PADEREWSKI so Peter must be using the same concept when forming both beliefs. This is true on the Originalist account of concepts as well, since it seems implausible that there is more than one originating use of PADEREWSKI. Hence, originalists face problems accounting for Peter's being rational, since both of Peter's PADEREWSKI tokens are of the same type due to there only being one public concept PADEREWSKI:

According to originalism, there is just one public concept PADEREWSKI, which Peter exercises both when he forms the belief that Paderewski has musical talent, and when he forms the belief that Paderewski lacks musical talent. In the originalist framework, Peter has contradictory beliefs: apart from negation, the beliefs are made up of just the same concepts in the same position. The challenge is to explain how Peter can, nonetheless, be rational. (Sainsbury & Tye 2012, 113—114)

The originalist cannot account for the puzzle of Paderewski in the same way they accounted for the puzzle of Hesperus and Phosphorus. What, then, can they do?

Two possible solutions seem to arise: First, one may deny that rationality hinges directly on the type of the concepts involved in cognitive processing. Second, one may supplement the theory with further machinery in order to account for the cognitive role of concepts. Originalists choose the first option. In what follows, I will present this approach. In section 4, I will assess this account and argue that the suggested solution fails to account for the puzzle of Paderewski and also that this approach renders the Originalist account incapable of explaining cognitive significance in general. In section 5, I will propose a solution in line with the second alternative: I will suggest what I take to be a minimal addition to any account that takes concepts to be public and show how this will allow such theorists to give a unified account of the intrapersonal cognitive role of concepts and thoughts.

The Originalist account of the puzzle of Paderewski ultimately involves denying that rationality hinges directly on the type of the concepts involved in cognitive processing. The originalists propose an account of rationality according to which individuals can be wrong about the number of concepts they possess. A common view is that individuals have privileged access to their own conscious mental states, including their thoughts. That is, it seems intuitively plausible that someone is in a position to know whether their thoughts are of the same type or not. Intuitively, one can know which concepts one possesses and whether they are of the same type or not. The thesis of introspective knowledge of comparative concepts (IKCC) may be stated as follows (c.f. Sainsbury & Tye 2012, 92):

**IKCC:** When our faculty of introspection is working normally, we can know apriori via introspection with respect to any two present, occurrent thoughts whether they exercise the same or different concepts.

Originalists reject IKCC. They thus reject Kripke's claim that "Anyone [...] is in a position to notice and correct contradictory beliefs if he has them" (Kripke 1979, 122).

In the case of Peter in the Paderewski puzzle, Peter does not know that his belief that Paderewski has musical talent contains the same concept PADEREWSKI as his belief that Paderewski lacks musical talent; he thinks his two thoughts contain different concepts. This is what accounts for Peter being rational in holding contradictory beliefs:

Peter thinks, at the rally, that he has learned a new concept PADEREWSKI, different from the PADEREWSKI concept he had earlier acquired at the concert. He is wrong. But this is a reasonable mistake, it is one that makes it reasonable for him to believe that his thoughts are not contradictory. He reasonably believes that it is not the case that the one thought consists in the other embedded in a concept of negation. (Sainsbury & Tye 2012, 134)

Since Peter has false beliefs concerning the identity of his deployed concept tokens, "it's rationally required of Peter to believe that the thoughts do not contradict" (Ibid., 135). This is so, even though Peter's two thoughts do in fact constitute a contradiction.

In what follows I will argue that this does not solve the puzzle of Paderewski. One of the questions raised by the puzzle concerns how Peter's tokens of PADEREWSKI may play distinct roles in cognition despite being of the same public type. The proposed solution does not give a satisfactory explanation of how this can be the case. Finally, I argue that the abandoning of IKCC renders the Originalist account of the cognitive role of concepts in terms of sameness and difference of concept type untenable in general.

#### 4. Originalism Fails to Account for Cognitive Significance

In what follows, I will argue that there are several problems with the Originalist approach to the puzzle of Paderewski. First, I will argue that the account involves over-intellectualization, and that this makes it unlikely that Originalism can provide a general account of the sort of cognitive processing involved in the puzzle of Paderewski. Second, I will argue that the account fails to explain how certain of Peter's beliefs concerning Paderewski are positively coordinated while others are not. Two concepts being positively coordinated, recall, just means that co-reference is manifest and as a result the concept tokens play the same role in cognition. Finally, I will argue that abandoning IKCC renders the originalist account of the cognitive role of concepts untenable in general.

##### *4.1. Over-Intellectualization*

The originalists explain the possibility of someone having contradictory first-order beliefs in terms of false second-order beliefs regarding the sameness or difference of constituent concepts of their first-order beliefs. In the case of Peter, he has a false second-order belief to the effect that the singular concepts deployed in his beliefs that Paderewski has musical talent and that Paderewski lacks musical talent are distinct. Although Peter is in fact deploying the same concept in both thoughts, Peter does not know this.

The initial problem with this account is that people rarely form higher-order beliefs about the nature of the concepts of their first-order beliefs. It is, for instance, unlikely that Peter consciously forms the belief that the two tokens of PADEREWSKI are distinct concepts.

Sainsbury & Tye foreshadows this worry and say that the sort of higher-order beliefs in question need not be explicitly formed:

One's introspective awareness that one is thinking that *p*, need not itself be manifest in consciousness by an occurrent thought [...] to

count as introspective knowledge, a belief need not itself be manifest in consciousness by an occurrent thought. For instance, one might introspectively know that one is thinking that water is a liquid on occasions in which the only occurrent thought one is having is that water is a liquid. (Sainsbury & Tye 2012, 90 n. 1)

It seems, then, that the originalist can respond to the initial worry by saying that Peter need not explicitly form the relevant higher-order belief, but instead that it is sufficient for his rationality that he implicitly believes the proposition.

However, even though it is not required that individuals explicitly form the relevant higher-order beliefs, they must still have the disposition to do so. One's disposition to believe something depends partly on the concepts one possesses. For instance, if someone lacks the concept CAT, this individual does not at that very moment have the disposition to form any beliefs involving the concept CAT. Some have argued that young children under the age of four lack the concept BELIEF (c.f. Gopnik 1993, Perner 1991). It is, in general, far from obvious that young children can entertain complex thoughts involving concepts such as BELIEF, CONCEPT etc. Such children would not have the disposition to form higher-order beliefs concerning their first-order beliefs or those beliefs' conceptual constituents. Even so, it seems that such children could fall victim to the sort of confusion involved in the Paderewski case. That is, they could have first-order beliefs deploying the same public concept without recognizing that the beliefs concern the same referent. If individuals may be in Paderewski-like situations without having the ability to form higher-order beliefs we cannot account for their rationality by appeal to false higher-order beliefs.

The originalists might object that they do not require the individual to actually form such higher-order beliefs to the effect that their first-order beliefs contain distinct concepts. Rather, the mere lack of a belief – explicit or implicit – to the effect that PADEREWSKI (figuring in belief 1) is of the same type as PADEREWSKI (figuring in belief 2) is enough for an individual such as Peter to be rational. Young children may then count as rational even in a Paderewski-like case, as long as they have no higher-order beliefs regarding the sameness of the

conceptual constituents of their first-order beliefs. This account, however, over-generalizes. If the absence of such higher-order beliefs is enough for a person to have negatively coordinated beliefs, small children would never have thoughts that are positively coordinated, since the absence of higher-order beliefs secures negative coordination. This is obviously wrong; it is clear that young children may have positively coordinated thoughts. For instance, they can trade on identity in making basic inferences about things in their environment. Hence, this move is untenable.

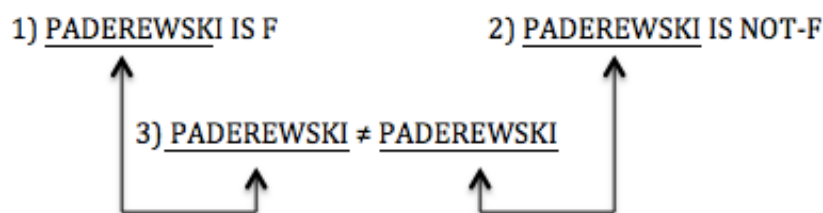
Whether or not young children have concepts such as BELIEF and CONCEPT is, of course, an empirical question. However, even if it should turn out that young children actually do possess such concepts, there are further and more serious problems with the Originalist account of the puzzle of Paderewski.

#### *4.2. The Problem of Coordination*

It seems that, according to the Originalists, Peter being rational depends on him making (or being disposed to make) something like the following inference:

- 1) PADEREWSKI HAS MUSICAL TALENT
- 2) PADEREWSKI DOES NOT HAVE MUSICAL TALENT
- 3) PADEREWSKI  $\neq$  PADEREWSKI
- 4) BELIEF (1) AND (2) ARE NOT CONTRADICTORY

The worry, then, is that is highly puzzling how (3) can do the necessary job in the inference. In order for (3) to do the necessary job, the two tokens of PADEREWSKI figuring in the belief must be positively coordinated with the tokens of PADEREWSKI in (1) and (2) respectively. We may visualize it thus (where the arrows represent positive coordination, while an absence of arrows represents negative coordination):



But how is the Originalist able to account for this? According to the Originalist, all four tokens of PADEREWSKI are of the same type. Still, somehow Peter ends up having only two pairs of positively coordinated concepts.

Appealing to identity judgments (or in this case, a judgment to the effect that there is no identity) in accounting for the cognitive role of concepts is notoriously problematic. Campbell (1994) addresses this when he considers accounting for coordination in terms of implicit or suppressed identity premises:

If this view were correct, we would also need to make sure that the uses of [PADEREWSKI] in the suppressed premise are linked with the uses of [PADEREWSKI] in the explicit premises, and we would need further suppressed premises to secure these connections. The problem recurs, and we are embarking on a regress. (Campbell 1994, 75)<sup>10</sup>

The originalists face the same problem in accounting for the inference from (1—4): If sameness of type does not guarantee positive coordination, how do we account for the positive coordination relations between the tokens of PADEREWSKI in (3) with the tokens of PADEREWSKI in (1) and (2) respectively? Appealing to Peter's higher-order belief concerning the identity (and lack thereof) of the constituent concepts in his first-order beliefs would lead to a vicious regress.

Hence, by giving up on the idea that sameness of concept type is transparent to thinkers the Originalists thereby give up on the idea that sameness or difference in concept types can account for coordination. As a result, the Originalists fail to

<sup>10</sup> Campbell uses 'Hesperus' as an example rather than 'Paderewski'.

account for Peter's rational processes. In the next section, I argue that the abandonment of IKCC has more far-reaching consequences than those concerning the puzzle of Paderewski. I argue that abandoning IKCC renders the Originalist account of coordination in terms of sameness and difference of concept type untenable in general.

#### *4.3. Originalists Fail to Account for the Fregean Data*

Let's return to the puzzle of Hesperus and Phosphorus. In this puzzle we want to know why it is the case that the members of the pair [HESPERUS, HESPERUS] are positively coordinated while the members of the pair [HESPERUS, PHOSPHORUS] are negatively coordinated. As we saw in 3.1, the Originalists' solution was to say that in the former case the two concept tokens are of the same type due to them having the same originating use, whereas in the latter case, the two concept tokens are of distinct types due to having distinct historical origins.

However, as we have seen, if IKCC is false, this means that two concepts that are of the same type may nonetheless play different cognitive roles. In contrast, two concepts that are positively coordinated will, necessarily, play the same role in cognition. If this were not the case, coordination would not be the sort of thing that could account for the Fregean data. That is, if coordination relations come apart from the cognitive role of concepts, it would be possible for individuals to have two concept tokens that were positively coordinated but that nonetheless played different roles in cognition. We would then have to introduce further notions in order to explain how someone could fail to recognize the co-reference of positively coordinated concepts. If sameness or difference in concept types is not transparent to thinkers (or, more specifically, their cognitive computational systems), coordination and concept types come apart, and as a result coordination cannot be accounted for in terms of sameness or difference in concepts. We are left with the following question: If sameness of type does not guarantee that two concept tokens are positively coordinated, why, in this specific case, do the two tokens of HESPERUS end up being positively



coordinated?

It is of little help when the Originalist claims that

The mechanism that leads to treating things as the same or different is a pretty reliable one, and it is essential to reasoning. One lesson of the puzzle of Paderewski is that the mechanism is not infallible. (Sainsbury & Tye 2012, 136)

Peter's cognitive mechanisms are normal, i.e. not malfunctioning. We want to explain why some of his thoughts concerning Paderewski are positively coordinated, i.e. manifestly co-referential, whereas other of his beliefs concerning Paderewski are negatively coordinated, meaning that the co-reference is not manifest to Peter. We still need to answer the question of why the sameness is recognized by the cognitive system in some cases while not in others. If the cognitive system can be wrong about the sameness or difference in types of concepts entertained in thought, such sameness or difference cannot explain rational reasoning. This is the case even in non-Paderewski style cases, since there is no principled distinction between the cognitive workings of someone in a Paderewski case and someone like the Ancient Babylonians. The only difference between the two cases is that in the former only one person is unaware of the co-reference of the relevant concept tokens, while in the latter an entire community lacks such knowledge. This, however, cannot make a difference to the intrapersonal cognitive mechanisms of single individuals.

When discussing the puzzle of Paderewski, the Originalists spell out the story of Peter in some more detail. They say that pragmatic facts concerning the circumstances of how he acquires the concepts as well as other beliefs Peter may have, play a part in how he is rational in holding contradictory beliefs. Experiencing the same object twice does not guarantee that the sameness of reference is recognized; this much is established by the puzzle of Hesperus and Phosphorus.

Further, Sainsbury & Tye consider a related puzzle, namely the puzzle about Paul who believes that cats have tails and also that chats have tails, but who doesn't

know that cats are chats. We may assume that CAT and CHAT is the same concept with a common origin. In this case, the Originalists say that

the explanation must be linguistic. In using the word “cat”, Paul, before his “discovery”, wrongly takes himself to exercise a different concept from the concept he exercises using the word “chat”.  
(Sainsbury & Tye 2012, 128)

Ultimately, then, the explanation of coordination seems not to rest on sameness or difference in concept types, but rather in whether or not the thinker *takes* the concepts to be the same. Whether or not an individual takes two concept tokens to be the same does not depend on whether or not it is actually the same concept, but rather it seems to fully depend on pragmatic facts about the specific circumstances in which a thinker acquires the concept.

I think this insight is ultimately correct. Coordination in thought does not map onto sameness or difference of public concept types. Instead, I suggest, it is a primitive fact about how the cognitive system stores and computes information. In the next section I will suggest a minimal addition to the Originalist framework – or more generally, any account that takes concepts to be public – that allows such theories to account for the interpersonal explanatory job traditionally assigned to concepts, such as accounting for communication, as well as coordination in thought.

### 5. Positive Proposal: Vehicle Relationism

I propose that the cognitive role of concepts is not to be accounted for in terms of sameness or difference in types of concept tokens, but rather it is to be accounted for in terms of primitive representational relations. The relations being primitive means that they do not hold in virtue of intrinsic representational features of concepts and thoughts. Intrinsic representational features are those properties of a concept that can be stated without reference to

another representation.<sup>11</sup>

The framework I have in mind is that of *Vehicle Relationism* (c.f. Chapter 1 and 2 in this theses). According to Vehicle Relationism, intrapersonal coordination is accounted for in terms of *pointer relations* and *mental tags*. I use the notion of ‘pointer relations’ to denote the sort of primitive relations that account for the cognitive role of mental representations. A pointer is to be understood as a relation that holds between representational vehicles. Whenever this relation holds the relata are positively coordinated. Importantly, the pointer relations do not reduce to sameness of intrinsic representational features of the relata. That is, it is not possible to determine whether or not two representational vehicles are positively coordinated merely by looking at the properties of the two vehicles in isolation. According to Vehicle Relationism two representational vehicles may share the exact same intrinsic properties – including the semantic content they express – and still fail to be positively coordinated. Further, two mental representations that are not connected by pointers are negatively coordinated.

I use the notion of ‘mental tags’ to denote those representational vehicles that are capable of entering into pointer relations. Mental tags and pointer relations are the building blocks of thoughts. The central claim relevant to our purposes in this chapter is that, when it comes to explaining coordination there is nothing of interest to be learned from looking at the intrinsic features of the tags – only relational features are explanatorily interesting when it comes to explaining coordination in thought.

What I propose is that the Fregean data that appeared to impose the Fregean Constraint on concept individuation are not to be accounted for in term of sameness or difference in types of concepts and thoughts, but rather, all we need to account for the data is a relational difference between the pairs of concepts [HESPERUS, HESPERUS] and [HESPERUS, PHOSPHORUS].<sup>12</sup>

---

<sup>11</sup> C.f. Gray (2017, 4).

<sup>12</sup> For a thorough explanation of how Vehicle Relationism accounts for the puzzle of Hesperus and Phosphorus as well as the puzzle of Paderewski see Chapter 2, section 3.

Heck (2012) argues for the same conclusion. They consider the difference in cognitive role of (1) and (2):

- 1) SAMUEL CLEMENS HAS DIED
- 2) MARK TWAIN HAS DIED

The two singular representations, CLEMENS and TWAIN are co-referential. Nonetheless they may play different roles in cognition. Heck illustrates this by considering a case involving an individual, Fred, who when he comes to believe (1) gets devastated, whereas he does not get equally saddened when coming to believe (2). Heck argues that

1. What distinguishes the belief that Clemens has died from the belief that Twain has died is nothing intensional. In particular, these beliefs have the same content.
2. If we are to be able to explain Fred's behavior in cognitive terms, there must be some difference between these beliefs that plays a role in psychological explanation. But no *intrinsic* difference between these beliefs plays that role. The explanatorily relevant difference is an extrinsic, relational one. It concerns how these beliefs are related to other of Fred's beliefs. (Heck 2012,144-145)

Heck focuses on the consequences of the Fregean data for intentional explanation and in stating intentional laws. They argue that for this specific purpose we only need the coordination relations (what they call *formal relations*) and as such, "the sorts of formal relations among beliefs that must be mentioned may be treated as psychologically primitive: We can make reference *directly* to these relations in giving intentional explanations and in stating intentional laws" (Ibid., 157). The key claim of Vehicle Relationism is that such relations obtain as a result of primitive relational features of mental representations.

The originalists themselves point to the explanatory importance of the sameness or difference in relational properties:

[One] role for concepts in cognition is that their sameness and difference generate relational effects: the pattern exemplified by the thought that Hesperus is Hesperus is distinct from the pattern exemplified by the thought that Hesperus is Phosphorus. (Sainsbury & Tye 2012, 53)

However, as we have seen they take such relational effects to hold in virtue of intrinsic representational features, namely concept tokens being of the same type or not:

In [the thought that Hesperus is Hesperus], a single concept is used twice. In the [thought that Hesperus is Phosphorus], two concepts are each used once. This may have an impact on informativeness. (Ibid.)

I think this is misguided, and that this is the reason why the Originalists fail to account for the cognitive role of concepts and thoughts. Publicly sharable concepts are too coarse-grained to account for the cognitive role of mental representations. Instead, we should follow Heck in taking the relational properties as primitive and account for the Fregean data in terms of such relations rather than sameness or difference in concept types. Heck's general observation is, however, compatible with coordination relations ultimately supervening on intrinsic representational features of concepts, such as sameness or difference in types (c.f. Heck 2012, 159). I suggest, however, that we should go one step further and say that such relations hold in virtue of primitive relations obtaining at the level of representational vehicle, such as pointers.<sup>13</sup>

There are several virtues of adopting this view of intrapersonal coordination. One such virtue is that concept types do nothing of explanatory interest when it comes to intrapersonal cognitive workings. The relevant rational capacities of individuals are fully accounted for in terms of primitive relational properties of mental representations. This means that the Fregean Constraint is no longer a constraint on concept individuation, since there is no longer a need to say that concepts that play distinct roles in cognition differ in anything beyond their

---

<sup>13</sup> For a motivation for making this further step, see chapter 2, section 4.

relational properties. The data that impose the Fregean Constraint in the first place are accounted for in such a way that there's no constraint on concept individuation that requires the sort of fine-grained individuation conditions usually taken to be required due to the supposed work done by sameness or difference in concepts in intrapersonal coordination.

This leaves us with the Publicity Constraint on concept individuation. What I suggest, then, is that we individuate concepts in a coarse-grained manner so as to account for interpersonal explanatory tasks such as communication. In other words, if sameness or difference in concepts does not explain intrapersonal cases, we avoid the Fregean Constraint and are thus free to say that the nature of concepts is such that they can be shared by distinct individuals.

A further question, then, concerns the exact nature of concepts. The nature of concepts must be such that they explain communication. Note that we still cannot say that concepts are individuated by their referents, since this renders the individuation conditions too coarse-grained to explain communication. That is, we still want to make a distinction between a case in which two distinct individuals believe that Hesperus is a planet and cases where one subject believes that Hesperus is a planet and the other that Phosphorus is a planet. Another option would be to individuate concepts by way of their senses, or the like. Note that by giving up on the claim that a difference in cognitive role is explained in terms of a difference in concepts we thereby also abandon Frege's initial motivation for introducing senses.

A further option is to individuate concepts not in terms of their semantic contents, but rather in terms of something non-semantic. Originalism provides one such alternative. That is, although Originalist concepts cannot explain intrapersonal cognitive processing, they can explain interpersonal cases, such as communication. HESPERUS and PHOSPHORUS having distinct origins (and thus being of distinct types) would explain the difference between two thinkers both believing that Hesperus is visible and a case in which one believes that Hesperus is visible while the other believes that Phosphorus is visible. At the same time, there is only one public concept PADEREWSKI, which would explain why Peter

could possibly successfully communicate with non-confused individuals about Paderewski. For instance, if Peter says “Paderewski has musical talent” to a non-confused individual, it seems like a case of successful communication if the hearer forms the appropriate thought deploying the public concept PADEREWSKI. Likewise, if Peter says “Paderewski is a politician”, it seems that he could also communicate successfully with non-confused individuals if they interpret the utterance by deploying their concept PADEREWSKI.<sup>14</sup>

The notion of pointer relations is a minimal addition to Originalism – or more generally, any theory that takes concepts to be public. At the same time it is explanatorily powerful. One could maintain that concepts are individuated in accordance with the Publicity Constraint, and that the relevant interpersonal phenomena, such as communication, is explained in terms of sameness or difference in concept types. At the same time, one would have to reject that coordination in thought is to be explained in terms of sameness or difference in such concepts. Hence, supplementing any theory that provides a good explanation of the public aspect of concepts with pointer relations has the virtue of enabling the framework to account for both the intrapersonal and the interpersonal consideration traditionally assigned to concepts.<sup>15</sup>

In recent years other views that make a similar move have been proposed. Prosser (2018) argues that by combining traditional views of the individuation of concepts with a view that posits primitive epistemic relations we can do justice to both the data that support the Fregean Constraint as well as the data that support the Publicity Constraint. In the next section I compare the proposed view to Prosser’s account.

---

<sup>14</sup> There might, of course, arise some confusion at some point during such a conversation. If Peter says both that Paderewski has musical talent and that Paderewski does not have musical talent, the hearer will most likely think that Peter is irrational or she might question what Peter is trying to communicate. Such confusion might, however, arise in any conversation in which what someone says conflicts with our beliefs about the speaker or the general states of affairs. If, for instance, you tell me that the earth is flat, I might think that you are crazy or that you were not really saying what it seemed like you were saying. This kind of confusion, however, is not explained in terms of a lack of coordination between our thoughts.

<sup>15</sup> Note that I do not intend this to be a defence of Originalism about publicly sharable concepts. There might be further problems for Originalism when it comes to explaining communication (or other tasks for concepts that require them to be sharable). There might be other coarse-grained ways of individuating concepts that are preferable to the Originalist way of doing it. If this is the case, we may combine such views with Vehicle Relationism and get the desired result.

## 6. An Alternative Approach

Thoughts can be positively coordinated within the mind of single individuals. They can also be positively coordinated across speakers. I have argued that one way to account for this is to take coordination in thought to be understood in terms of primitive relations, and that this would allow us to account for coordination between different individuals in terms of sameness or difference in concept types. In what follows, I will sketch an alternative approach, which takes the opposite route: It takes coordination across distinct individuals to be accounted for in terms of relational features of thought. This opens up the possibility of accounting for the intrapersonal cognitive data in terms of sameness or difference in concepts.

Positive coordination relations, as I have understood them here, are just those relations in virtue of which (putative) co-reference is manifest to the thinker(s). Positive coordination thus gives warrant for trading on identity. The view I want to sketch here is one that takes interpersonal coordination to hold in virtue of relational features of thoughts within the minds of distinct individuals. The kind of view I have in mind is one that claims that such relations cannot be reduced to sameness or difference in the type of the concepts and thoughts in question. There are some accounts in the literature that could potentially be compatible with this claim. For instance, Prosser (2018) holds that interpersonal coordination holds in virtue of epistemic relations between thoughts in the minds of distinct thinkers.

Roughly, Prosser's central claim is that singular concepts – or modes of presentations (MOPs) – are shared between distinct individuals just in case co-reference is transparent to thinkers in communication in such a way that they may trade on identity:

I suggest that in cases of transparent communication, where the speakers trade on identity rather than relying on interpretive



premises, the speakers think of the reference under the same MOP.  
(Prosser 2018, 8-9)

He accounts for such transparent communication in terms of individuals sharing a language and intending to use the words in this language in accordance with how others in their language community use these words. That is, on this account, deference to others plays an important role in linguistic communication. A precondition for transparent communication – i.e. the sort of communication where the interlocutors trade on identity – is that the interlocutors share words in this way.<sup>16</sup>

There is, in particular, one pertinent interesting feature of the epistemic relations Prosser has in mind. He says that

Strictly speaking the relation that we capture by saying that [S1] thinks of O under the same MOP as [S2] is not an identity relation but an intransitive transparency relation. (Ibid., 15)

Since the relevant relations are not transitive they cannot be further accounted for in terms of sameness or difference in concept types, since sameness of type is transitive.<sup>17</sup> Rather, on this view, interpersonal coordination is essentially accounted for in terms of relational features of thoughts. On this account, then, the observations that gave rise to the Publicity Constraint is accounted for not in terms of concept types, but rather in terms of relational features of concepts and thoughts. As a result, we no longer have to consider the Publicity Constraint as providing constraints on concept individuation.

This opens up the possibility of individuating concepts in a fine-grained way, in accordance with the Fregean Constraint. We could then explain the intrapersonal

---

<sup>16</sup> Prosser notes that “strictly speaking it is not shared words per se that facilitate interpersonal trading on identity; it is their coordinating role that matters” (Prosser 2018, 16).

<sup>17</sup> Note that this is not always the case in diachronic intrapersonal cases. For instance, if we construe singular concepts as mental files we can, for instance, imagine cases in which two distinct files merge into one file over time as a result of an identity judgement (c.f. Recanati 2016, Prosser 2018). In this case, the resulting file may be seen as a continuance of each of the previous files. Even though each of the early files are distinct from each other, they are in a sense the same as the merged file. Hence, transitivity fails.

cognitive data in terms of sameness or difference in types of concepts and thoughts. Prosser's general view would, for instance, be compatible with Fodor's (1975, 2008) Language of Thought Hypothesis, according to which coordination is accounted for in terms of sameness or difference in Mentalese syntactic symbols. It is also compatible with the mental file account of coordination (c.f. Perry 1980, Recanati 2012, 2016), according to which coordination is accounted for in terms of sameness or difference in such files.

It is beyond the scope of this paper to determine which framework is preferable. The main aim of this section has been to compare the proposed view to another similar possibility and in doing so illustrate the novelty of the proposed account of coordination. Note also that the framework of Vehicle Relationism is not committed to accounting for interpersonal coordination in terms of sameness or difference in types of concepts. This will only be a viable option if there is a theory of concepts that allows us to give such an account. Another possibility is to combine Vehicle Relationism with an account of interpersonal coordination in terms of epistemic relations along the way of Prosser (2018). In short, any account that does well in explaining interpersonal coordination may adopt Vehicle Relationism about intrapersonal coordination (and vice versa).

## 7. Conclusion

We have seen that there are two plausible constraints on concept individuation: The Publicity Constraint and the Fregean Constraint. Both constraints seem intuitively plausible on their own, but jointly they appear to be incompatible. The reason is that the Publicity Constraint requires concepts to be individuated in a coarse-grained way, whereas the Fregean Constraint requires concepts to be individuated in a fine-grained fashion. In this paper I have assessed one view of concept individuation, namely Originalism, which promises to overcome such difficulties.

I have argued that Originalism about concepts fails to account for the cognitive role of concepts. The reason why Originalism fails to do so is that the puzzle of

Paderewski forces the Originalist to say that sameness or difference in concept type is not transparent to thinkers. As a consequence, concept types and coordination come apart. This observation may be generalized: Any account that takes coordination to be accounted for in terms of sameness or difference in concept types and at the same time holds that such sameness or difference in concepts is non-transparent to thinkers fails to account for coordination. This is because coordination relations are, necessarily, transparent to thinkers.

I have suggested that by making a minimal addition to theories that take concepts to be public, such as Originalism, we may overcome such worries. The suggested idea is to take coordination in thought to be accounted for in terms of primitive relations holding between the vehicles of representation. We thus avoid the Fregean Constraint on concept individuation. We may then individuate concepts coarsely in such a way that we satisfy the Publicity Constraint. The resulting framework has the virtue of accounting for both intrapersonal coordination as well as interpersonal coordination.

## Staying on Topic: The Continuance-of-Topic Relation is Non-Transitive

*Semantic drift occurs whenever a term changes meaning over time. Sometimes such change is accompanied by a change in topic, while other times the topic remains stable. It may seem natural to draw a clear distinction between cases of semantic drift in which the topic is preserved and cases in which it is not. In this paper I argue that things are not always this easy. I argue that the continuance-of-topic relation is non-transitive. By doing this, I reveal a structural problem with popular accounts of topic stability according to which continuance of topic is accounted for in terms of identity relations, be it identity of concepts (c.f. Sawyer 2018, Richard forthcoming) or identity of meaning (Ball forthcoming).*

### 1. Introduction

Semantic drift occurs whenever a word changes meaning over time. This is a common phenomenon. Consider for instance the word 'clue'. It used to mean a ball of thread (of the sort that would help guide someone out of a labyrinth, for instance), whereas today it means something like evidence or information that helps solve a given task. The meaning of the word has thus changed. In this case,

the semantic drift has resulted in a change of topic: If someone were to utter

1. The police officers are looking for clues

in the 17<sup>th</sup> century, they would be talking about something different than someone uttering (1) today. The 17<sup>th</sup>-century person would be saying that the police are looking for balls of threads, whereas a speaker uttering (1) today would be saying that the officers are looking for any information that may help them solve a crime. The two utterances would have different truth conditions. Importantly, the difference in truth-condition of an utterance of (1) in the 17<sup>th</sup> century and today is not due to any relevant changes in the nature of balls of threads or the like. Rather, the change is purely due to a change of the semantic properties of 'clue'. As a result of this change, the 17<sup>th</sup>-century person and the current-day individual are not talking about the same thing when uttering (1). This is because there has been a change in topic. Semantic drift often results in a change in topic in this way.

Some cases of semantic drift, however, preserve topic. Take for instance the word 'fish'. When people in the 15<sup>th</sup> century thought and talked about fish, their utterances of this word were used to pick out whales as well as other aquatic animals. Today, however, we no longer use the term 'fish' to pick out whales. If we assume that a community's use of a term should be reflected in the meaning of that term, this would mean that the word 'fish' when uttered in the 15<sup>th</sup> century meant something different than what the word means today. Even so, this case seems to be different from the case of 'clue' in an interesting way. For instance, if someone in the 15<sup>th</sup> century uttered

2. Fish are aquatic animals

it seems correct to say that they were, at least in some way, saying the same as someone uttering the same sentence today. Further, it seems fine for someone to say, "People used to think that whales are fish". Contrast this with someone today saying, "People used to think that only balls of threads could be clues". The latter seems at best highly misleading. The relevant difference between the two cases is this: the case of 'clue' involves a change in topic, whereas in the case of

'fish' there has been stability of topic. This is so despite the semantic drift.<sup>1</sup>

In recent years the notions of semantic drift and stability of topic have gained currency as a result of the increased popularity of revisionary projects such as *conceptual engineering*. Conceptual engineering is the process of assessing and improving our words and other representational devices (c.f. Cappelen 2018). Revisionary projects involve changing semantic aspects of representational devices, but in many cases we still want to say that the topic of such representational devices has remained the same. The issue, however, is more far-reaching: Without stability of topic despite semantic change, it is hard to see how scientists can engage with pre-theoretic questions, given how scientific endeavour often requires a sharpening of our representational devices. Further, scientific progress would be hard, since many of the debates that drive science forward involve disagreement on the exact semantic properties of central terms. Even so, much of the work done on topic stability in recent years is found within the debate on conceptual engineering.

There has been a tendency to make a sharp distinction between, on the one hand, cases of semantic drift where the topic is preserved and, on the other, cases of semantic drift where there is a change in topic (e.g. Sawyer 2018). When it comes to accounting for continuance of topic the dominant strategy has been to appeal to identity of some sort. Some hold that sameness of topic is to be explained in terms of sameness of concepts (e.g. Sawyer 2018, Richard forthcoming). According to this view, continuance of topic is compatible with semantic drift. Others argue that stability of topics is explained in terms of stability of meaning, but that meaning does not change in the way we typically assume (c.f. Ball forthcoming). These accounts fall under a general approach,

---

<sup>1</sup> I use the case of 'fish' to illustrate the issue in an intuitive way. I base this example on Sainsbury's (2014) use of the term in order to illustrate semantic drift and continuity of topic (see Sawyer (2018) for this interpretation of Sainsbury's puzzle). Later in this paper I will discuss a view that denies the possibility of continuance of topic despite semantic drift: According to Ball (forthcoming) continuance of topic is explained in terms of sameness of meaning. Although I accept the possibility of topic stability despite semantic drift, I do not wish to beg the question against Ball's view. I will say more about Ball's view on topic stability and how it relates to the argument in this paper in section (3.2.). What's important is that proponents of Ball's model would agree that in the case of 'clue' there has been a change in topic, whereas in the case of 'fish' there has been continuance of topic.

which I will call *The Identity Approach to Topic Continuity*. This includes all accounts that explain continuity of topics in terms of identity relations.

In this paper I argue that all accounts that try to explain stability of topics in terms of identity relations fail. I show that we cannot draw a clear distinction between the cases of semantic drift that preserve topics and those that do not. This is because we have cases that in some sense fall into both categories. Some cases of semantic drift preserve topic along the way but still there is discontinuity of topic at the beginning and end of the evolutionary chain. That is, I will show that the continuance-of-topic relation is non-transitive. As a result, the identity approach fails, since identity is transitive.

The structure of the paper is this. In section 2, I say more about continuance of topics and why we ought to think that there can be such stability in cases of semantic drift. In doing this, I will make clear what explanatory work the notion of continuance of topics is supposed to do. In section 3, I elaborate on the different versions of the Identity Approach mentioned above. In section 4, I argue that the continuance-of-topic relation is non-transitive. Finally, in section 5, I address Cappelen's (2018) account of stability of topics. According to this view, *the Contestation Theory of the Limits of Revision*, we cannot give necessary and sufficient conditions for sameness of topics. In short, the view seems to be that there is not too much we can say in general about the nature of the continuance-of-topic relation; topic continuity is established on a case-to-case basis. The aim of this paper is not to propose a novel theory of how to account for continuance-of-topic relations. For all I know, Cappelen might be right that there is not an easy answer to this question. However, I will argue that the non-transitivity of topic continuity puts some restrictions on our account of continuance of topics. I will argue that the contestation theory can be correct only as long as we accept certain views about *samesaying*.

## 2. Continuance of Topic and Semantic Drift

I take the possibility of topic stability despite semantic drift as a data point. However, some philosophers, most famously Strawson (1963), has raised objections that may challenge this assumption. Why, one may ask, is it so important to think that there can be stability of topic despite semantic drift? Wouldn't it be easier to just assume that all semantic drift necessitates a change in topic? There are several reasons why we really ought to insist on stability of topics. In this section I will focus on what I take to be the central reasons why we need to do so. First, we want to explain how there can be continuity of inquiry despite semantic drift. Second, it seems highly plausible that individuals who disagree on the correct use of a certain representational device are having a substantial disagreement and not merely a verbal dispute. As we'll see, both of these explanatory tasks seem to require stability of topics. I take this to be the main explanatory tasks for which we need the notion of topic continuity. I will consider and elaborate on each explanatory task in turn.

### *2.1. Continuity of Inquiry*

In philosophical and scientific theorizing we often take pre-theoretic notions as outset and improve on such notions in order to have a more fruitful discussion. The process of improving our pre-theoretical notions for the sake of clarity in scientific theorizing is explicitly defended by Carnap in his (1950), where he develops his account of *explication*. In this process, an imprecise ordinary expression is transformed into an exact expression suitable for scientific endeavour. The central example given by Carnap involves the relation between the pre-theoretic notions of 'warm' (*classificatory*) and 'warmer than' (*comparative*) and the more scientifically useful notion of 'temperature' (*quantitative*). He says that,

A *quantitative* concept serves to describe something with the help of numerical values (e.g., temperature) [...] Quantitative concepts are no



doubt the most effective instruments in the scientific arsenal. (Carnap 1950, 8-9)

Carnap thus held a view according to which our representational devices may undergo changes in order to yield a more fruitful scientific investigation. In short, scientists are often trying to answer questions asked at the pre-theoretic outset by deploying new or improved representational devices.

As a response to Carnapian explication, Strawson raises the following problem:

Typical philosophical problems about the concepts used in non-scientific discourse cannot be solved by laying down the rules of exact and fruitful concepts in science. To do this last is not to solve the typical philosophical problem, but to change the subject. (Strawson 1963, 505)<sup>2</sup>

Further, he says that,

However much or little [Carnap's method of explication] is the right means of getting an idea into shape for use in the formal or empirical sciences, it seems *prima facie* evident that to offer formal [explications] of key terms of scientific theories to one who seeks philosophical illumination of essential concepts of non-scientific discourse, is to do something utterly irrelevant— is a sheer misunderstanding, like offering a text-book on physiology to someone who says (with a sigh) that he wished he understood the workings of the human heart. (Ibid., 504–505)

That is, if we change our representational devices in such ways, we seem to no longer be concerned with the questions we were asking at the pre-theoretic outset. If this is correct, philosophers are not really in the business of investigating pre-theoretic questions, but rather they seem to be working exclusively within their own bubble.

---

<sup>2</sup> I take 'topic' and 'subject matter' to be synonymous. A change in subject thus means a change in topic. I use these terms interchangeably throughout the paper.

Although Strawson's criticism is specifically directed towards Carnapian explication, the challenge can be generalized to all processes that involve significant changes to our representational devices but where we still want to say that people are talking about the same subject matter. With the increased popularity of revisionary projects such as conceptual engineering, Strawson's challenge is as relevant as ever. The generalized Strawsonian challenge is this: All projects that involve changing the semantics of key terms change the topic of the investigation, resulting in people talking past each other and philosophers providing solutions to problems other than the ones originally posed. This, of course, is by no means specific to philosophy: As progress and new discoveries are made, key terms change within all sciences.<sup>3</sup> If all semantic drift results in a change in subject matter, how can we ever make progress? The view that semantic drift necessitates a change in topic thus threatens scientific progress in general and should be avoided if possible.

The general response to the Strawsonian challenge is to simply deny that a change in semantics necessitates a change in topic. One reason to think that this response is correct is that two words uttered may differ in their overall semantic value but still we judge them to concern the same subject matter. Cappelen (2018, 110) suggests one such example: Imagine two speakers, A and B, both uttering the sentence "Serena is really smart". Even if we assume that the two speaker's contexts differ in many ways, including their exact assumptions about smartness so that A and B's respective contexts fix somewhat different extension and intension for 'smart', there are contexts in which it is true to say that A and B both said that Serena is really smart, or simply that A and B said the same thing about Serena. That is, Strawson's insistence that semantic change necessitates a change in topic does not fit well with our everyday assessment of whether or not two utterances concern the same topic. In general, without further argument it seems highly unlikely that just because our ways of describing the world become more precise and fruitful, it follows that we are not talking about and trying to understand the same subject matter.

---

<sup>3</sup> See for instance Brigandt (2010) for an overview of how the word 'gene' has changed within biology. Other examples include the terms 'species' and 'planet' (c.f. Bokulich 2014).

One of the central explanatory tasks for the notion of continuance of topic is thus to account for the possibility of continuance of inquiry in cases of semantic drift. If there is continuity of inquiry despite semantic drift of key terms, this is because there is topic continuity. I now turn to the second reason for thinking that we may have stability of topic despite semantic drift, namely the need for an account of substantial disagreement.

## *2.2. Disagreement*

Let's return to the case of 'fish' discussed at the outset, and consider a case posed by Sainsbury (2014). Sainsbury tells the story of James Maurice, who in 1818 was the inspector of fish oil and tax collector of New York City. Maurice filed a lawsuit against a man named Samuel Judd, accusing him of labelling barrels of whale oil wrongly in order to avoid paying fees. Judd responded by saying that he was not to pay the fees for fish oil, since the oil in his barrels were whale oil, and whales are not fish. In front of the jury each side invited experts on the topic in order to convince the jury that they were right. Anatomists argued that whales are not fish while merchants and seafarers argued that whales are indeed fish. In the end, the jury ruled in favour of Maurice, ruling that whales are fish. Sainsbury describes the process thus:

The parties offered significant reasons for their opinions. The victorious side reasoned from the premises that sea-creatures are fish, and whales are sea-creatures. The first premise is enshrined in much popular opinion, including creation stories according to which God made the creatures of the air (birds), the creatures of the land (beasts) and the creatures of the sea (fish). The losing side argued on the basis of significant dissimilarities between fish and whales: fish oxygenate using gills, reproduce by laying eggs and have true fins, whereas whales oxygenate using lungs, reproduce by suckling live-born young, and their fin-like appendages are differently structured from the fins of, say, a shark. There was no dispute about these facts,

only about their relevance. Both parties agreed that whales are lung-using, air-breathing mammals and that they are sea-creatures. (Sainsbury 2014, 3)

If we assume that the meaning of a word in a community is determined by how it is used in that community, there seem to be two distinct meanings associated with the term 'fish'.<sup>4</sup> The lawsuit in question appears to take place at a point in history where the meaning is gradually changing. The merchants and seafarers (as well as most people up until that point in history) were using the word 'fish' to pick out whales as well as other aquatic animals, and the anatomists (as well as most people after this point in time) used the word so as to not include whales. Even so, the different agents clearly seem to be having a substantial disagreement when one of the parties were asserting "whales are fish" and the other party denied this. If their words had different meanings, how could we account for such disagreement? This is Sainsbury's puzzle:

we may presume that the ancients used 'fish' with a meaning on which 'Whales are fish' was true. The meaning of their word 'fish' would align with their use: they used it to include whales, and that fits its meaning. Then there is only a verbal difference between the disputants in *Maurice vs. Judd*. 'Whales are fish', understood in the ancient way, is true, but, understood in the modern way, it is false. This is inconsistent with the fact that the debate was substantive, and not merely verbal. (Ibid., 4)

The way Sainsbury cashes out the notion of a substantive disagreement is that such disagreement requires agreement in meaning (2014, 4). There must be a proposition that one part affirms and the other denies. This is what creates the puzzle.

---

<sup>4</sup> As mentioned in a previous footnote, the possibility of topic continuity despite semantic drift is controversial. Assuming that there is continuity of topic in the case 'fish', Ball (forthcoming) would deny that this case involves a difference in meaning of the relevant term. I follow Sainsbury (2014) and Sawyer (2018) in taking cases like the one above to involve a semantic change. Importantly, my criticism of Ball's view (section 4 below) does not depend on such details.

However, Sawyer (2018) argues that there need not be sameness in meaning in order to account for such disagreement. Rather, what accounts for this sort of disagreement is continuity of topic (I return to Sawyer's proposed solution to Sainsbury's puzzle in more detail in section 3.1. below). If this is correct, we need the notion of topic continuity to account for the possibility of substantial disagreement despite semantic drift. Thus understood, accounting for substantial disagreement is one of the central explanatory tasks for the notion of topic continuity.

Based on the considerations in this section, I understand the continuance-of-topic relation to be such that if there is continuity of inquiry and possibility of substantial disagreement, the continuance-of-topic relation obtains. Put differently, I will call whatever relation accounts for the possibility of continuity of inquiry and substantial disagreement in cases such as the ones looked at above, *the continuity-of-topic relation*. A natural question to ask, then, is in virtue of what does this relation obtain? In the next section I will put forth two suggestions as to how continuance of topic obtains found in the literature. Both suggestions are examples of what I call *The Identity Approach to Continuance of Topics*.

### 3. The Identity Approach

In this section I focus on two general strategies for accounting for the continuance-of-topic relation found in the literature. These strategies explain continuance of topic in terms of identity of some feature. The first strategy I consider is to account for stability of topics in terms of sameness of concepts. The general approach is to say that continuance of topics is more coarse-grained than the intension and extension of a term. The second strategy is to ultimately deny that there can be stability of topics despite a change in meaning. Proponents of this view do however hold that there is stability of topics in the kind of cases discussed in the previous section of the paper. Taking this as a starting point, they deny that the cases involve a change in meaning. Both of

these general strategies are examples of The Identity Approach to continuance of topics. I'll look at each strategy in turn.

### *3.1. Identity of Concepts*

The first strategy is advocated by Sawyer (2018), who argues that the stability of topic through a linguistic change is to be explained in terms of stability of concepts. Such concepts are understood as mental particulars that are individuated in terms of their topics. That is, concepts are individuated in terms of relations to objective properties or objects that are independent of our conception of such properties or objects:

It is the concept expressed by a term that determines its subject matter. This is, in effect, a consequence of the fact that the subject matter itself enters into the individuation conditions of the relevant concept. (Sawyer 2018, 137)

As a result, a term can express the same concept at different times despite being associated with different conceptions.

Sawyer thus makes a clear distinction between the explanatory role of concepts and linguistic meaning. On this framework, the linguistic meaning of a term is determined by use:

I suggest that the linguistic meaning of a term at a time be understood as the characterization of the relevant subject matter that members of the linguistic community would settle on at that time were they to reach reflective equilibrium in the context of a dialectic. (Ibid., 130)

So understood, a term may change its meaning while still being used to express the same concept. On this view, then, sameness of concept (and thereby continuance of topic) is more coarse-grained than sameness of meaning.

One thing to note is that Sawyer makes a clear distinction between cases of linguistic change that involve a change in topic and those that do not:

From a philosophical perspective, meaning shift falls into two broad categories: cases that are accompanied by a corresponding change in subject matter; and cases that are not (Ibid., 127—128).

I take this to be a widespread view – after all, either there is a change or there is not a change. I will, however, challenge this assumption in section 3 of this paper.

Sawyer uses her framework to account for Sainsbury's puzzle concerning the semantic difference between an utterance of "whales are fish" in the 15<sup>th</sup> century and now (addressed in the previous section). In doing this, she focuses on the semantic change in 'whale' rather than 'fish', but I take it that this doesn't make an essential difference to this line of response.<sup>5</sup> People in the 15<sup>th</sup> century would settle on a characterization of whales according to which the animals were described as fish. Today, however, our linguistic community would not characterize whales as fish. Since, on Sawyer's framework, meaning is determined by the agreed upon characterization of the relevant subject matter in a community, the meaning of the term 'whale' has changed. This explains why, in the 15<sup>th</sup> century the sentence "whales are fish" was generally regarded as true, whereas today we would regard the sentence as false.

Despite this, Sawyer says, there is stability of topic, since there has been a stability of concepts associated with the term 'whale'. Throughout the linguistic change of 'whale', the individuals in the linguistic communities stood in relation to the same kind of animal. As a result, there has been stability of topic despite the semantic drift. This is why, despite the difference in meaning between the anatomists' and the fishermen's use of the term 'whale', they could be said to have a substantial disagreement. Sawyer thus ascribes the same explanatory role to continuance of topic as I did in section (2.2).

---

<sup>5</sup> If someone were to object to Sawyer's response to Sainsbury's puzzle on the grounds that the case of 'whale' is not analogous to the case of 'fish', that's fine. However, this is not the objection I want to raise in this paper.

To sum up, Sawyer takes continuance of topic to be explained in terms of sameness in concepts. A similar line of thought is found in Richard (forthcoming). He agrees that there might be continuance of topics despite semantic drift and that such stability is explained in terms of concepts. His view differs from that of Sawyer in what he takes the nature of concepts to be. On Richard's view, concepts are to be understood as 'stretched out' objects that persist over time. The continuance of topics is explained in terms of words expressing concept time slices that belong to the same stretched out concept. On this model, two concept time slices will concern the same topic if they belong to the same stretched out concept.<sup>6</sup>

Although the two views differ with respect to their views on the nature of concepts, the general strategy is the same: Continuance of topics is to be explained in terms of sameness of concepts – be it concept tokens being of the same type or concept time slices belonging to the same stretched out concept. A consequence of both views is that the continuance of topic relation must be transitive, since sameness (i.e. identity) is a transitive relation.

### *3.2. Identity of Meaning*

Whereas the views considered in the former paragraphs hold that sameness of topic is more coarse-grained than sameness of meaning, the view I'll consider next denies this. However, it does so in a novel and relevant way. Instead of saying that revising our representational devices leads to a change in topic due to a change in meaning, Ball (forthcoming) argues that such revisions do not result in a change of meaning, and as a result the topic remains the same.

The key claim is that, despite appearances to the contrary, cases such as 'fish' and 'whale' do not involve a change in meaning. Instead, the meaning of a word is fixed by our final (accepted) stipulation about that word's meaning. Ball says that "on the subject continuity model that I advocate, there is no new meaning; a

---

<sup>6</sup> C.f. Cappelen's (2018, 144–146) construal of Richard's (forthcoming) *Species Model of Concepts*.



successful stipulation fixes the meaning of the word as it was used all along” (Ball forthcoming, 19). So for instance, if we assume that there will be no further revisions made to the word ‘fish’, the meaning of the word was always so that it excluded whales. When the merchants and fishermen argued in court that whales are fish, they were using the word with the same meaning as we are today, although they were not aware of the full meaning of the word. The same goes for the ancient language community in which no one ever doubted that whales are fish. People in such communities were simply wrong, and when they uttered (5):

##### 5. Whales are fish

they were saying something false.

Note that proponents of this view would not be happy with the way I’ve set out the discussion thus far. In presenting the debate, I’ve been assuming the possibility of semantic drift in certain cases of topic-continuity. On Ball’s view, there would not really be semantic drift in such cases. If there is sameness of topic, the meaning of a word is the same. Although I have been assuming the possibility of continuance of topics despite semantic drift, I need not insist on this. What I would have to insist on is stability of topics in cases such as the ones considered above where there *appears* to be a change in meaning. Ball would agree that there is continuance of topics in such cases. What he would dispute is that such cases involve semantic drift. Instead he would say that the meaning has stayed constant, but that this meaning is determined by the last agreed upon stipulations. Importantly, the argument in the next part of the paper holds regardless of whether or not one accepts the possibility of semantic drift in cases of topic continuity.

A key claim of Ball’s subject continuity model is that revisionary endeavours often maintain a stability of topic. In fact, I take this to be one of the motivating observations for the view. Ball considers the alternative view, the view that revisionary analysis necessarily involves a change in subject matter. This is the *subject-change view* of revisionary analysis. He argues against this view partly on

the basis of what he calls *The Argument Argument*. Consider two individuals who disagree on whether or not same-sex couples should be allowed to marry. The person who thinks same-sex couples should not be allowed to marry thinks so on the grounds that they also believe that the purpose of marriage is to produce children. In a debate between the two individuals, this latter person utters (6):

6. The purpose of marriage is to produce children

Now Balls says that in general,

As a matter of methodology, we should look for an interpretation of these arguments and our responses to them that makes sense of what we are doing. (Ball forthcoming, 6)

This suggests, he says, the following requirement on our interpretation of a debate about a revisionary analysis:

It must give parties to the debate reasonably good epistemic status with respect to the things they say. In many typical cases, this will mean that the parties to the debate are saying things that they know, or at least justifiably believe, to be true. [...]

It must make assertions relevant to the debate. When a party to the debate makes an assertion, that assertion should serve some purpose: for example, by giving evidence for her position or evidence against her opponent's position. (Ibid.)

If we assume that debates concerning how to revise a concept or term necessarily involve a change of subject matter, it is hard to make sense of the behaviour of the person uttering (6). If she uses the term in her own preferred way, let's call this marriage<sub>1</sub>, she is saying something trivial and not relevant to the discussion since that use of the term 'marriage' would by definition exclude same-sex couples. As a matter of fact, if this were the case, her opponent could very well accept her utterance, knowing that the statement has nothing to do with her own position. At the same time, if the speaker intended to use the word

the way her opponent prefers to use it, call this marriage<sup>2</sup>, she should consider (6) as an obvious falsehood, but she doesn't. Ball concludes that the subject-change view cannot make sense of the person's utterance of (6). As a result,

the subject-change view simply fails to make sense of the way we argue for and against revisionary analyses; and even in those cases where the subject-change view has a partial story to tell, it often fails to explain the whole phenomenon. (Ball forthcoming, 10)

Ball's argument does at least partly depend on the claim that stability of topics is needed to make sense of these kinds of debates.

Ball's subject continuity model accommodates the view that revisionary endeavours often maintain a stability of topic. It does so, however, by denying that such endeavours may involve a change in meaning of words in cases where the topic remains stable. In this sense, Ball advocates a clear distinction between cases where we have a change in topics and cases where the topic remains the same. In the next section I will argue that this is a mistake. I will show that topic continuity is more complex; there are cases of topic change that happen so gradually that no point in history marks a discontinuity with immediate previous uses of the word. I argue that both of the general approaches looked at in this section of the paper fail to account for these observations.

#### 4. The Continuance-of-Topic Relation is Non-Transitive

The aim of this section is to establish that the continuance-of-topic relation is non-transitive. I will do this through a case study, namely the evolutionary change of the term 'meat'. In the 17<sup>th</sup> century 'meat' meant anything edible. That is, 'meat' used to pick out food in general. Today, however, the term picks out animal flesh only. It is clear that the case of 'meat' involves a change in topic. Sawyer mentions this case as an example of topic change that is relatively straightforward, since the case involves a change in meaning as well as in topic:

The extensions of terms such as ‘meat’ [...] change over time because different entities satisfy their descriptive meanings at different times. Meaning shift of this first kind is also accompanied by a change in subject matter. What makes meaning shift of this first kind relatively straightforward is the fact that the intuitive change in subject matter of the relevant terms is marked, in each case, by the change in extension that results from the change in linguistic meaning. (Sawyer 2018, 133)

However, as we’ll see below, if the case of ‘meat’ is portrayed according to Sainsbury & Tye’s (2012, 46) stipulations, the case becomes anything but straightforward. In fact, I will argue that the case of ‘meat’ is the best illustration of the complexity of the continuance-of-topic relation.

#### *4.1. The case of ‘Meat’*

Let’s start by acknowledging the fact that there has been a change in meaning of the term ‘meat’ from the 17<sup>th</sup> century until today.<sup>7</sup> Sainsbury & Tye focus on linguistic meaning in general rather than topics, but, as Sawyer points out, it is clear that there is a discontinuity of topic as well. Now, throughout the years the meaning of ‘meat’ gradually drifted until one day the word was used to pick out animal flesh only. Importantly, at no point did anyone make a decision to abandon the old use; everyone deferred to other uses in their language community when using the word. Even so, a semantic drift happened along the way. But at no point was the drift significant enough for anyone to pick up on it as it was happening.

We may construe the evolution of the meaning change as a timeline starting at  $t_1$  when the word picked out food in general, and then there were a number of

---

<sup>7</sup> Again, Ball (forthcoming) would deny this claim. He might say that the term ‘meat’ had the same meaning all along, or that there are different terms with distinct meanings involved in this story. I base the presentation of the story of ‘meat’ on Sainsbury & Tye’s (2012) construal of the case, according to which there was a change in meaning. Importantly, the argument to follow does not depend on details about the meaning of the term.

microscopic steps,  $t_2, t_3, t_4 \dots t_n$ , along the way until today, which we may call  $t_{2018}$ . Now, although there has been a change in topics from  $t_1$  until  $t_{2018}$ , there was no discontinuity of topic at any of the small steps  $t_n—t_{n+1}$  along the way.

To see why this is the case, consider the following scenario: Someone in the 17<sup>th</sup> century wants to explore the harms and benefits of eating food, as opposed to not eating anything at all. This person would ask the question:

5. What are the health harms and benefits of eating meat?

Further, let's say that someone today wants to investigate the health harms and benefits of eating meat, i.e. animal flesh. This person would also ask question (5). Even though it is stipulated that the two individuals are using the same terms when formulating their questions, there seems to be an obvious discontinuity of inquiry between their two projects. If this is the case, there cannot be sameness of topic between 'meat' in the 17<sup>th</sup> century and now, since sameness of topic would guarantee continuity of inquiry in this case.

Further, we could imagine that at any given time in the history of the evolution of 'meat' there was an individual asking (5). Since the semantic drift happens so gradually without any point in history marking a discontinuity with the immediate previous use, it is the case that at any time it is true that there was continuity of inquiry when someone asked question (5) at  $t_n$  and when someone asked the same question at  $t_{n+1}$ . Hence, since there is continuity of inquiry, it must be the case that there is continuity of topic at any time  $t_n$  and  $t_{n+1}$ .

Now, consider another case: Someone in the 17<sup>th</sup> century asserts (6):

6. Vegetarians do not eat meat

This utterance would be false, since vegetarians do eat food, and the person making the statement seems to be confused about what a vegetarian is. Many years later, in 2018, someone asserts (6) again. Due to the change of reference of 'meat' this utterance is true. Knowing this, it is hard to imagine people being happy with someone today disquotationally reporting on the 17<sup>th</sup>-century

speech act. That is, if we were to say that someone in the 17<sup>th</sup> century said that vegetarians do not eat meat, this would be highly misleading and, arguably, false. This indicates, again, that there must be a difference in topic between the utterance of (6) in the 17<sup>th</sup> century and an utterance of the very same sentence today. Even so, it seems perfectly fine to say that at any time,  $t_n$ , during the semantic evolution of the term ‘meat’ it is fine to disquotationally report an utterance of (6) at  $t_{n-1}$ .

Furthermore, in the case above, the person sincerely uttering (6) in the 17<sup>th</sup> century does not have a substantial disagreement with the person sincerely negating the same sentence in 2018. To make this clear, imagine a further case in which our 17<sup>th</sup>-century individual is no longer confused about what a vegetarian is and utters, “Vegetarians *do* eat meat”. This is true, since vegetarians eat food. Contrast this with someone today saying, “Vegetarians do *not* eat meat”. This is also true, due to the semantic drift. Importantly, this seems like a case in which we would not be happy to say that the two individuals disagree about what vegetarians eat. To make it even more vivid, add a time machine to the story: The person from the 17<sup>th</sup> century travels to our time where she meets our contemporary speaker. Naturally, they soon enter into a chat about the dietary requirements of vegetarians. At first it might appear to them and people around that they are having a disagreement, but this is only superficially so and it will quickly be discovered that the interlocutors are merely talking past each other. No one would say that they actually disagree about what vegetarians do or do not eat. Hence, there is no continuity of topic in the beginning and end of the chain of deference. Still, just like before, it seems perfectly coherent to stipulate that at any time,  $t_n$ , a sincere utterance of “vegetarians do eat meat” would constitute a disagreement with someone sincerely uttering “vegetarians do not eat meat” at  $t_{n+1}$ . This, again, shows that there is continuity of topics at any time  $t_n$  and  $t_{n+1}$ .

What makes the case of ‘meat’ interesting is that at no point during the history of the term does it make sense to say that there was discontinuity of topic along the way. From one day to the next there was always continuity of topic in utterances involving the word ‘meat’. However, as we have seen, at some point during the

semantic drift there came a time in which there was no longer continuity of topic with the original case, where the term ‘meat’ was first introduced. The continuance-of-topic relation, then, must be non-transitive: It is possible that two utterances, *a* and *b*, deploy continuity of topic and that utterance *b* and *c* also deploy such continuity, and at the same time that utterance *a* and *c* do not deploy continuity of topic.

If the argument in this section is correct, any view that accounts for continuance of topic in terms of identity relations fails, since identity is transitive. Let’s look at each of the frameworks looked at in section 2 to see, in some more detail, why this is the case.

#### 4.2. *The Identity Approach: Concepts*

Let’s return to the identity of concept view: Proponents of this account would have to say that ‘meat’ as used in the 17<sup>th</sup> century expresses a different concept than ‘meat’ used today, since the terms concern different topics. However, in order to make sense of the fact that every step along the way preserves topic, they would have to say that it is the same concept expressed at every time  $t_n$  and  $t_{n+1}$ . These two claims are inconsistent and so the concept identity view lacks the resources to account for the non-transitivity of topics in cases such as ‘meat’.

Proponents of Sawyer’s view might want to object that I have not given a correct characterization of the ‘meat’ case. On the view we’re considering, concepts are individuated in terms of relations to individuals and objective properties in the external world. Take, for instance, the case of ‘whale’:

the term ‘whale’ expresses the concept *whale* in the actual world both at  $t_1$  and at  $t_2$  in virtue of the fact that the linguistic community at  $t_1$  and the linguistic community at  $t_2$  both stand in the requisite relation to whales. (Sawyer 2018, 135)

One might want to argue against the non-transitivity of topic continuity by saying that in the case of ‘meat’, there are actually distinct concepts involved at every minor step  $t_n$  and  $t_{n+1}$  since there is a slight difference in the group of entities one is related to. This would, however, be problematic. If sameness of concepts is to account for continuance of topics and continuance of topic is needed to account for continuance of inquiry and substantial disagreement, we fail to explain how someone at  $t_n$  and  $t_{n+1}$  could be said to have a substantial disagreement or how there could be continuity of topic between these points in time in the case of ‘meat’. As we have seen, Sawyer explicitly uses her account of topic continuity to give an account of substantial disagreement. In order to account for the potential of having substantial disagreements between any two microscopic steps along the evolutionary history of the term ‘meat’, she would therefore have to say that it is the same concept. I thus take it that the sameness of concept view fails to account for continuance of topics as a result of the continuance-of-topic relation being non-transitive.

#### *4.3. The Identity Approach: The Subject Continuity Model*

Ball’s subject continuity model faces similar worries. In order to explain how there could be a continuity of topic of ‘meat’ at every step  $t_n$  and  $t_{n+1}$  proponents of the subject continuity model would have to say that the tokens of the term have the same meaning. Remember that, on this view, meaning is fixed by the final agreed upon stipulations. Since, by assumption, tokens of ‘meat’ at  $t_{2018}$  and tokens of the same term at  $t_{2018-1}$  deploy topic continuity, this means that they will have the same meaning. We could repeat this for every minor step all the way back to  $t_1$  in the 17<sup>th</sup> century. However, this would have the consequence that the word ‘meat’ had the same meaning at  $t_1$  as in  $t_{2018}$ , since sameness of meaning is transitive. However, if this is correct, there must be continuity of topic between utterances involving ‘meat’ then and now. This, as we have seen, is clearly not the case, since there could not be substantial disagreement or continuity of inquiry involving the term ‘meat’ then and now. Hence, the subject continuity model lacks the resources to explain how there can be continuity of



topics at every step through the evolutionary chain but discontinuity of topics at the beginning and end of the chain.

Given Ball's commitment to the idea that continuity of topics is needed to explain individuals' argumentative strategies it is hard to see how he could deny this. Let's say someone, at  $t_1$  asserts, "vegetarians eat meat" in a letter. Let's call the writer of the letter S1 and the receiver S2. By the time the letter reaches S2, we're at  $t_2$ . At this point, S2 writes back saying "as a matter of fact, vegetarians do *not* eat meat". In order to make sense of this argumentative strategy, we should grant that S2 is talking about the same topic as S1 when using the term 'meat'. S2 does not disagree with S1's use of 'meat'. She merely thinks vegetarians do not eat food. Analogous cases could be repeated between every minor step along the way from  $t_1$  until  $t_{2018}$ . By Ball's own *Argument Argument* it seems to follow that we must say that the interlocutors are talking about the same subject in order to make sense of their argumentative strategies.

Now, imagine if our current-day time-traveller were to engage in a discussion with S1. Assuming that she is aware of the difference between how people used the term 'meat' at  $t_1$  and today, she would simply grant S1's statement, and not respond by saying "vegetarians do *not* eat meat". The reason is that she knows that such an assertion would not be relevant to the conversation. Instead, she should accept S1's utterance, "Vegetarians eat meat", as not conflicting with her own belief as well as her potential linguistic response. Note that I am not assuming anything about the actual meaning of the word 'meat' in this case. Rather, I'm hypothesising about the interlocutor's argumentative strategies. If, as Ball argues, sameness or difference in subject matter is to account for such argumentative strategies, it seems to follow that we must say that S1 and S2018 (i.e. our current-day individual) are not talking about the same subject matter in the case of 'meat' in order to explain why they do not argue in the way we would expect had their utterances concerned the same subject matter.

As a result of this, it seems that in order to make sense of these argumentative strategies we must say that there is continuance of topic at every minor step in the evolutionary history of 'meat', but not continuance of topic at the beginning

and end of the chain. Hence, Ball's Subject Continuity Model fails to account for the non-transitivity of the continuance-of-topic relation. In the next sub-section I will consider one more possible objection on behalf of the proponents of the Identity Approach.

#### *4.4. The Sorites Objection*

The argument presented against the Identity View may be stated thus:

P1: If the Identity View is correct, then it is the case that continuance-of-topic relation is transitive

P2: It is not the case that the continuance-of-topic relation is transitive

C: The Identity View is not correct.

The argument is valid. P1 is uncontroversial. Proponents of the Identity View would then have to deny P2. I have given good reasons to think that P2 is true. The Identity Theorist would have to deny that the sort of case I've presented (i.e. the case of 'meat') is possible. It's important to note that I've presented the case of 'meat' in a theory-neutral way. That is, in presenting the case I have not assumed anything uncontroversial about the continuance-of-topic relation. It is uncontroversial that the continuance-of-topic relation is necessary for continuity of inquiry and the possibility of substantial disagreement. The case of 'meat' illustrates that we may have continuity of inquiry and substantial disagreement at any minor step through history, but that there nonetheless could be discontinuity of inquiry and no substantial disagreement at the beginning and end of the evolutionary chain. On the basis of this, I have concluded that the continuity-of-topic relation is non-transitive.

The objection to my argument that I wish to consider in this section is one according to which the case is to be understood as a Sorites case. If this is correct, the Identity Theorist could argue that the case of 'meat' does not pose a problem for their view, since in general, hardly anyone thinks that the response to standard Sorites cases is to deny that the relation in question is transitive.

There are various competing accounts of the sort of vagueness involved in Sorites cases, and the Identity Theorist may simply adopt whatever seems to be the best account on the market. To take an example, the Identity Theorist may follow Williamson (1994) and argue that our ignorance about thresholds in borderline cases does not mean that there cannot be such thresholds. If this is the case, they may claim that even though it seems to the individuals involved as though there is continuity of inquiry and substantial disagreement at any point  $t_n$  and  $t_{n+1}$ , there must be some point in history where this is not the case. That is, there must be a point at which people are simply mistaken about whether or not they have a substantial disagreement or whether there is continuity of inquiry.

This response comes with a considerable theoretical cost for the Identity Theorist. Let's assume that the cut-off point in question happened at a given time,  $t_n$ . The Identity Theorist would then have to say that there was continuity of inquiry and possibility of substantial disagreement between  $t_{n-1}$  and  $t_n$ , but not between  $t_n$  and  $t_{n+1}$ . She will have to insist that this is the case despite the fact that the way in which the case is set up, the amount of change between any two points will be symmetric. She would then have to explain why it is that people had the behavioural patterns appropriate to there being continuity of topic despite the lack thereof. For instance, on Ball's view, we would have to account for people's argumentative strategies in terms of something else than continuity of topics. If someone at  $t_n$  (right before the cut-off point) uttered "vegetarians eat meat" and someone at  $t_{n+1}$  (right after the cut-off point) took themselves to disagree and as a response uttered "vegetarians do not eat meat", we cannot explain this argumentative strategy in terms of continuity of topic. This would conflict with Ball's claim that continuity of topic is needed in order to account for agents' rational argumentative strategies. This seems like a serious cost for a theory that is committed to explaining individuals' argumentative strategies in terms of continuity of topics. Further, the Identity Theorist would not only have to explain why the people in the story thought that they were having substantial disagreements between  $t_n$  and  $t_{n+1}$  but also why it seems so sensible to us, the spectators, to ascribe substantial disagreement to such individuals without the appeal to continuity of topics.

Due to the theoretical cost, I take this route to be explanatorily unattractive. But what are the alternatives? In the next paragraph I will consider another possible – and potentially more promising – account of the continuance-of-topic relation, namely Cappelen’s (2018) *Contestation Theory*. I will argue that the non-transitivity of topic continuity puts some important restrictions on this view.

### 5. The Contestation Theory of The Limits of Revision

Cappelen agrees that continuance of topic does not track sameness of semantic properties. He takes continuance of topic to be more coarse-grained than sameness of extension and intension:

Two sentences with different semantic contents (where semantic content is understood as, at least, having the same extension and intension) can be used to say the same thing, or to talk or be about the same topic (Cappelen 2018, 108).

He also agrees that the continuance-of-topic relation is complex. In fact, the key claim of his *Contestation Theory* is that we cannot give necessary and sufficient conditions for continuance of topics. Instead, such continuity (or the lack thereof) is to be established on a case-to-case basis.

Prima facie this seems to fit well with the observations made in the previous part of the paper: The continuance-of-topic relation is too complex for us to make a clear distinction between cases in which the topic has been stable and cases where the topic has changed. Giving necessary and sufficient conditions in terms of identity of some sort – be it identity of concepts or identity of meaning – fails to accommodate the intricate nature of the continuance-of-topic relation. The non-transitivity of topics does, however, present a need to qualify some of the claims of the *Contestation Theory*.

Although Cappelen maintains that one cannot give necessary and sufficient conditions for stability of topics, he relates the notion of topics to the notion of

samesaying: “if A and B said the same, then they are talking about the same thing and their topic is the same” (Cappelen 2018, 113). This claim figures into his response to the Strawsonian challenge, discussed in section 2.1. The response “appeals to data about when we correctly describe people as *having said the same thing*” (Ibid., 107). The argument runs as follows:

The first step in this argument points out that A and B can samesay each other using a sentence ‘Fa’, even though the extension of ‘F’ in A’s speech differs from the extension of ‘F’ in B’s speech. [...] The second step in the argument says that if samesaying is possible despite differences in extension, then so is ‘talking about the same topic’. Sameness of topic goes hand in hand with samesaying (Ibid., 107-108).

The general idea, I take it, is that whenever it is appropriate to say that two people are samesayers they are talking about the same topic. This gives us some guidelines as to what to check for when we want to assess (on a case-to-case basis) whether or not two utterances concern the same topic. If, for instance, it seems appropriate to say that the scientist and the layperson are saying the same when using a particular term, they are indeed talking about the same topic, and so we have a response to the Strawsonian challenge. This view, when combined with the observation that the continuance-of-topic relation is non-transitive, has some interesting consequences for the notion of samesaying.

We may distinguish between two broad views on the nature of samesaying. On the one hand, we have views according to which samesaying is a matter of similarity of contents. On the other, we have views according to which samesaying requires identity of (at least some aspects of) content. The latter view is defended by Cappelen and Lepore in their (2007). Here they argue against the view that the samesaying locution does not require content identity across contexts. Cappelen and Lepore’s main argument against the view that samesaying is a matter of similarity runs as follows:

Similarity is not transitive. Transitivity is, however, built into the 'said that' locution. Consider first (T1):

**(T1)** If A said what B said, and B said what C said, then A said what C said.

If, however, A said something similar to what B said, and B said something similar to what C said, it simply doesn't follow that A said something similar to what C said. (T1), according to the similarity theory, could be false. Since T1 can't be false, the similarity theory fails. 'A said that p' simply does not mean the same as 'A said something similar to p'. [...] as far as we can tell, this is about as close to a conclusive objection to the view that 'A said that p' means the same as 'A said something similar to p' that any will ever come up with. So, we could just end the critical discussion here. (Cappelen & Lepore 2007, 124)

However, samesaying can only truly go 'hand in hand' with the continuance-of-topic relation as long as the former is also a non-transitive relation. Consequently, the view of samesaying advocated by Cappelen and Lepore (2007) is not available to proponents of Cappelen's Contestation Theory. If we were to combine the claim that samesaying and continuance of topics go hand in hand with the view that samesaying requires identity of content this would result in another version of the Identity View of topic continuity. I've shown that the Identity View of topic continuity fails, and as a result the Contestation Theory is a plausible account of topic stability only as long as it is combined with a similarity account of samesaying. By adopting a similarity view of samesaying, the contestation theory avoids the problems posed for the Identity View.

The similarity account of samesaying renders the samesaying relation non-transitive, and so it fits well with how we would ascribe samesaying in the case of 'meat': At every minor step along the way of the historical evolution of 'meat' we would be happy to ascribe samesaying to an utterance of "meat is edible" at  $t_n$

and an utterance of the same sentence at  $t_{n+1}$ . Still, we would not ascribe samesaying to an utterance of the sentence in the 17<sup>th</sup> century and today.

## 6. Conclusion

I have argued that the continuance-of-topic relation is non-transitive. By doing this, I have revealed a structural problem with the Identity Account of stability of topics: Topic stability cannot be accounted for in terms of transitive relations, such as identity.

Further, the findings of this paper show why it is wrong to draw a clear distinction between cases in which topic is preserved and cases in which topic is not preserved. In cases such as 'meat' the topic has changed since the term was first introduced. Even so there is a chain of topic-continuity between all minor evolutionary steps all the way back to the first use of the term.

The upshot of these findings is that the continuance-of-topic relation is a lot more complex than traditionally assumed. More investigation into exactly how complex the relation is is needed. It's possible that the factors that determine whether or not the relation obtains are so complex that we cannot give necessary and sufficient conditions for the continuance-of-topic relation to obtain. Although there is something unsatisfactory with this sort of account, at least it is preferable to the Identity View of topic stability in that it does not conflict with the non-transitive nature of the continuance-of-topic relation.

## Bibliography

- Aydede, M. (2010). "The Language of Thought Hypothesis", In E. N. Zalta, (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition).
- Ball, D. (Forthcoming). "Could Women be Analytically Oppressed", In A. Burgess, H. Cappelen, and D. Plunkett, (eds.), *Conceptual Ethics and Conceptual Engineering*. Oxford: Oxford University Press.
- Barber, A. (2000): "A Pragmatic Treatment of Simple Sentences", *Analysis* 60(4): 300–308.
- Bokulich, A. (2014). "Pluto and the 'Planet Problem': Folk Concepts and Natural Kinds in Astronomy", *Perspectives on Science* 22(4), 464—490.
- Brigandt, I. (2010). "The epistemic goal of a concept: accounting for the rationality of semantic change and variation", *Synthese* 77(1): 19—40.
- Burge, T. (1979). "Sinning Against Frege". Reprinted in his (2005): *Truth, Thought, Reason: Essays on Frege*, Oxford: Oxford University Press
- Campbell, J. (1987). "Is Sense Transparent?" *Proceedings of the Aristotelian Society* 88: 273—92.
- \_\_\_\_\_. (1994). *Past, Space and Self*. Cambridge, Mass.: MIT Press.
- \_\_\_\_\_. (2002). *Reference and Consciousness*, Oxford: Oxford University Press.
- \_\_\_\_\_. (2012). "Perceiving the Intended Model", In P. Machamer and A. Raftopoulos (eds.), *Perception, Realism and the Problem of Reference*: 96—122. Cambridge: Cambridge University Press.
- Caplan, B. (2007). "Millian Descriptivism", *Philosophical Studies* 133(2): 181—190.
- Cappelen, H. (2018). *Fixing Language*. Oxford: Oxford University Press.
- Cappelen, H. & Lepore, E. (2007). "Relevance Theory and Shared Content". In N. Burton-Roberts (Ed.), *Advances in Pragmatics*. Palgrave Macmillan.



- Carnap, R. (1950). *The Logical Foundations of Probability*. University of Chicago Press
- Carruthers, P. (1989). *Tractarian Semantics: Finding Sense in Wittgenstein's Tractatus*, Oxford: Blackwell.
- Crimmins, M. (1992). *Talk About Beliefs*, Cambridge, Mass.: MIT Press
- Crimmins, M & Perry, J. (1989). "The Prince and the Phone Booth: Reporting Puzzling Beliefs", *The Journal of Philosophy* 86(12): 685—711.
- Davidson, D. (1963). "Actions, Reasons and Causes", *Journal of Philosophy* 60(23): 685–700.
- Duhau, L. (2012). "The Myth of Concept Publicity", *Ideas y Valores* 61(148): 101—113.
- Evans, G. (1982). *The Varieties of Reference*, Oxford: Oxford University Press.
- Fiengo, R. & May, R. (1996). "Anaphora and Identity". In S. Lappin (ed.) *Handbook of Contemporary Semantic Theory*, 117–144. Oxford: Blackwell.
- \_\_\_\_\_. (1998). "Names and Expressions", *Journal of Philosophy* 95(8): 377–409.
- Fine, K. (2007). *Semantic Relationism*, Oxford: Blackwell.
- \_\_\_\_\_. (2010a). "Comments on Scott Soames' 'Coordination Problems'", *Philosophy and Phenomenological Research* 81(2): 475—484.
- \_\_\_\_\_. (2010b). "Reply to Lawlor's 'Varieties of Coreference'", *Philosophy and Phenomenological Research* 81(2): 495—501.
- Fodor, J. (1975). *The Language of Thought*. Cambridge, Mass.: Harvard University Press.
- \_\_\_\_\_. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, Mass.: MIT Press.
- \_\_\_\_\_. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.
- \_\_\_\_\_. 2008. *LOT 2: The Language of Thought Revisited*. Oxford: Oxford University Press.
- Fodor, J., & Pylyshyn, Z. (1988). "Connectionism and Cognitive Architecture: A Critical Analysis", *Cognition* 28(1-2): 3—71.

- Frege, G. (1892). "On Sense and Reference", In P. Geach & M. Black (eds.), *Translations from the Philosophical Writings of Gottlob Frege* (1960), Oxford: Blackwell.
- \_\_\_\_\_. (1980). *Philosophical and Mathematical Correspondence*, Chicago: University of Chicago Press.
- Gopnik, A. (1993). "How we know our minds: The illusion of first-person knowledge of intentionality", *Behavioral and Brain Sciences*, 16(1), 1-14.
- Gray, A. (2017). "Relational Approaches to Frege's Puzzle", *Philosophy Compass* 12(10).
- \_\_\_\_\_. (2018). "Indistinguishable Senses", *Noûs*, online first.
- Haslanger, S. (2000). "Gender and race: (what) are they? (what) do we want them to be?", *Noûs* 34(1), 31—55.
- Heck, R. K. (2002). "Do Demonstratives Have Senses?", *Philosophers' Imprint*, 2(2): 1—33.
- \_\_\_\_\_. (2012), "Solving Frege's puzzle", *Journal of Philosophy* 109(1-2): 132—174.
- Heim, I. (1982). *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. Dissertation, University of Massachusetts, Amherst.
- \_\_\_\_\_. "On the projection problem for presuppositions", In P. Portner & B. H. Partee (eds.), *Formal Semantics - the Essential Readings*, 249—260. Oxford: Blackwell.
- Hendriks, H. (2002) "Information Packaging: From Cards to Boxes", In K. van Deemter and R. Kibble (eds.) *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*, 1–33. Stanford, Calif.: CSLI.
- Kamp, H. (2015). "Using Proper Names as Intermediaries Between Labelled Entity Representations", *Erkenntnis* 80(2): 263—312.
- Kripke, S. 1979. "A Puzzle about Belief", In A. Margalit (ed.), *Meaning and Use*, 239–283. Reidel, Dordrecht.
- \_\_\_\_\_. (1980). *Naming and Necessity*. Cambridge, Mass.: Harvard University Press.
- Laurence, S. & Margolis, E. (2007). "The Ontology of Concepts – Abstract Objects or Mental Representations?", *Noûs*, 41(4): 561—593.

- Lawlor, K. (2002). "Memory, Anaphora, and Content Preservation", *Philosophical Studies* 109(2): 97—119.
- \_\_\_\_\_. (2010). "Varieties of Coreference", *Philosophy and Phenomenological Research* 81 (2): 485—495.
- Lockwood, M. (1971). "Identity and Reference". In M. Muniz (ed.) *Identity and Individuation*: 199—211. New York: New York University Press.
- Lycan, W. & Prinz, J. (2008). *Mind and Cognition: An Anthology*. 3rd edition. Oxford: Blackwell.
- Mates, B. (1952). "Synonymity", In L. Linsky (ed.) *Semantics and the Philosophy of Language*, 111—138. Urbana Ill.: University of Illinois Press.
- McDowell, J. (1984). "De Re Senses", *Philosophical Quarterly* 34(136): 283—294.
- \_\_\_\_\_. (1994). *Mind and World*. Cambridge, Mass.: Harvard University Press .
- Merricks, T. (2015). *Propositions*. Oxford: Oxford University Press.
- Mill, J. S. (1843). *A System of Logic*, London: Parker
- Millikan, R. (1997). "Images of Identity", *Mind* 106(423): 499-519.
- \_\_\_\_\_. (2011). "Losing the Word-Concept Tie." *Proceedings of the Aristotelian Society*, Supp. 111: 125—143.
- Ninan, D. (2015). "On Recanati's Mental Files", *Inquiry: An Interdisciplinary Journal of Philosophy* 58(4): 368—377.
- Onofri, A. (2016). "Two Constraints On A Theory Of Concepts", *dialectica* 70(1): 3—27.
- \_\_\_\_\_. (2018). "The Publicity of Thought", *The Philosophical Quarterly* 68(272): 521—541.
- Peacocke, C. (1992). *A Study of Concepts*, Cambridge, Mass.: MIT Press.
- Perner, J. (1991). *Learning, development, and conceptual change. Understanding the representational mind*. Cambridge, Mass.: MIT Press.
- Perner, J., Huemer, M., and Leahy, B. (2015). "Mental files and belief: A cognitive theory of how children represent belief and its intensionality", *Cognition* 145: 77—88.
- Perry, J. (1980). "A Problem about Continued Belief", *Pacific Philosophical Quarterly* 61: 317—332.

- Pinillos, A. (2011). "Coreference and Meaning". *Philosophical Studies* 154(2): 301—324.
- Plunkett, D. & Sundell, T. (2013). "Disagreement and the semantics of normative and evaluative terms", *Philosophers' Imprint*, 13(23): 1—37.
- Prinz, J. (2002). *Furnishing the Mind*, Cambridge, Mass.: MIT Press.
- Prosser, S. (2018). "Shared Modes of Presentation", *Mind & Language*, online first.
- Recanati, F. (2012). *Mental Files*. Oxford: Oxford University Press.
- \_\_\_\_\_. (2016). *Mental Files in Flux*. Oxford: Oxford University Press.
- Rey, G. (1994). "Concepts", In S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*: 185—193. Oxford: Blackwell.
- Richard, M. (1990). *Propositional Attitudes: An essay on thoughts and how we ascribe them*. Cambridge: Cambridge University Press.
- \_\_\_\_\_. (Forthcoming). *Meaning as Species*. Oxford: Oxford University Press.
- Russell, B. (1903). *The Principles of Mathematics Vol. I*, Cambridge: University Press.
- \_\_\_\_\_. (1910). "On the Nature of Truth and Falsehood", In *Philosophical Essays*. Cambridge: Cambridge University Press.
- \_\_\_\_\_. (1912). *The Problems of Philosophy*. Oxford: Oxford University Press.
- Sainsbury, M. (2014). "Fishy business", *Analysis* 74(1): 3–5.
- Sainsbury, M. & Tye, M. (2011). "An Originalist Theory of Concepts". *Proceedings of the Aristotelian Society*, Supp. Vol. 85: 101—24.
- \_\_\_\_\_. (2012) *Seven Puzzles of Thought*. Oxford: Oxford University Press.
- Sawyer, S. (2018). 'The Importance of Concepts'. *Proceedings of the Aristotelian Society*, 118(2), 127—147.
- Soames, S. (2002). *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*, New York, NY: Oxford University Press.
- \_\_\_\_\_. (2010). "Coordination Problems", *Philosophy and Phenomenological Research* 81(2): 464—474.
- Speaks, J. (2010). "Millian Descriptivism Defended", *Philosophical Studies* 149(2): 201—208.

Stanley, J. (2011). *Know How*. Oxford: Oxford University Press.

Strawson, P. F. (1963). "Carnap's views on conceptual systems versus natural languages in analytic philosophy". In P. Schilpp (Ed.), *The Philosophy of Rudolf Carnap*. 503—518. La Salle: Open Court.

\_\_\_\_\_. (1974). *Subject and Predicate in Logic and Grammar*. London: Methuen.

Vallduví, E. (1992). *The Informational Component*. New York: Garland.

Williamson, T. (1994). *Vagueness*. London and New York: Routledge.



