

# Cruel Intensions: An Essay on Intentional Identity and Intentional Attitudes



Alexander Sandgren

A thesis submitted for the degree of  
Doctor of Philosophy of the Australian National University  
(Submitted 8<sup>th</sup> of October 2015, passed without revisions on the 21<sup>st</sup> of April 2016)

This thesis is solely the work of its author. No part of it has previously been submitted for any degree, or is currently being submitted for any other degree. To the best of my knowledge, any help received in preparing this thesis, and all sources used, have been duly acknowledged.



---

ALEXANDER SANDGREN

8<sup>th</sup> of October 2015

# Acknowledgements

Firstly, I want to thank Daniel Nolan, the chair of my supervisory panel. Daniel has been a patient, constructive, and encouraging mentor and friend. I could not have asked for a better supervisor.

I would also like to thank the other members of my supervisory panel, David Chalmers, Daniel Stoljar, and especially Frank Jackson. In developing the ideas in this thesis, I have benefited greatly from engaging with Frank's work as well as from lively discussions with Frank.

In writing this thesis, I have benefitted greatly from discussions with members of the ANU philosophical community, both past and present. Special thanks are due to Rachael Briggs, Rachael Brown, Adrian Currie, Clare Due, Edward Elliott, Alan Hájek, Jessica Isserow, Robert Kirby, Erick Llamas, Donald Nordblom, Philip Pettit, Koji Tanaka, and Lachlan Umbers. I would also like to thank regular attendees of the ANU metaphysics reading group for enlightening and stimulating discussions.

I would also like to thank Ben Blumson, Scott Dixon, David Ripley, Pär Sundström, Caroline Torpe Touborg, Alexandra Varlakov, and Robbie Williams for intriguing discussions on everything from decision theory to the ontology of fictional objects.

I am grateful for the continuing support of my parents, Håkan and Robyn, and for their love and faith in me. Thanks are also due to my sisters, Hanna and Mia, who have long been sources of inspiration for me.

Finally, I would like to express my sincere gratitude to my wife, Deirdre. Thank you for your unwavering support, kindness, and friendship. I could not have done this without you.

# Abstract

Some intentional attitudes (beliefs, fears, desires, etc.) have a common focus in spite of there being no object at that focus. For example, two beliefs may be about the same witch even when there are no witches, different astronomers had beliefs directed at Vulcan, even though there is no such planet. This relation of having a common focus, whether or not there is an actual concrete object at that focus, is called *intentional identity*. In the first part of this thesis I develop a new theory of intentional identity, *the triangulation theory*, and argue that it has significant advantages over the extant theories of intentional identity in the literature. Empty attitudes (attitudes that are not, *prima facie*, about anything that exists) will serve as useful cases for testing theories of intentional identity.

In the second part, I put the theory developed in the first part to work. I use triangulation theoretic tools to shed light on other debates about intentional attitudes. Some issues to which intentional identity are relevant are the debate about the *content* of intentional attitudes, the issue of whether or not we need to appeal to external constraints on the content of intentional attitudes, how we should understand the agreement and disagreement of attitudes, how we should construe communication and how we ought to solve Kripke's puzzle about belief. The second part of this thesis also motivates a broadly internalist and individualistic approach to the content of intentional attitudes; it turns out that if we take a closer look at the narrowly construed psychological states of agents we find materials that allow us to make sense of phenomena usually associated with externalist constraints on the content of attitudes (such as causal constraints and eligibility constraints) in a new way.

# Contents

Chapter 1: Preliminaries .....	2
1.1. Intentional identity .....	2
1.2. Why we should care .....	4
1.3. What we are looking for in a theory of intentional identity .....	7
1.4. What the project is <i>not</i> .....	9
1.5. The plan.....	10
Chapter 2: Which Witch is Which? Exotic objects and Intentional Identity .....	12
2.1. Intentional identity and transparency .....	12
2.2. Exotic objects .....	15
2.3. Assignment.....	16
2.4. The hard cases .....	17
2.4.1. Abundance .....	18
2.4.2. Poverty.....	18
2.4.3. The natural suggestion.....	21
2.5. The problem .....	22
2.6. Responses .....	24
2.6.1. Resisting abundance, poverty or the natural suggestion.....	24
2.6.2. Distinctive creationist options? .....	25
2.6.3. Refusing the challenge.....	27
2.6.4. Hybridising .....	27
2.7. Conclusion.....	30
Chapter 3: More Theories of Intentional Identity .....	33
3.1. Alternatives .....	33
3.2. Dennett's theory .....	33
3.3. The lazy theory.....	34
3.4. King's theory.....	36
3.5. Manning's theory .....	36
3.6. Geach's theory and Sainsbury's theory.....	37

3.7. Glick’s theory and van Rooy’s theory.....	38
3.8. Complaint 1: Incompleteness .....	39
3.9. The causal link approach.....	41
3.10. Complaint 2: Which causal links?.....	44
3.11. Complaint 3: Causal links are not necessary for intentional identity.....	46
3.12. An extra challenge.....	47
3.13. Complaint 4: Non-uniformity .....	48
3.14. Complaint 5: Symmetry .....	48
3.15. Glick’s strategy .....	50
3.16. Conclusion.....	54
Chapter 4: A New Theory of Intentional Identity.....	56
4.1. The triangulation theory .....	56
4.2. Intentional distinctness.....	61
4.3. Disagreement.....	62
4.4. Flexibility .....	62
4.5. No appeal to exotica.....	64
4.6. The advantages of the triangulation theory .....	65
4.7. Extensions and qualifications.....	70
4.7.1. Attitudes with more than one focus.....	70
4.7.2. Dependent triangulation conditions.....	72
4.7.3. Other attitudes.....	73
4.7.4. Attitudes about kinds and properties .....	73
4.7.5. Triangulating with more than one agent.....	74
4.7.6. Symmetric intentional identity ascriptions .....	74
4.7.7. Ignorance of g-relatedness and mistaken beliefs about g-relatedness .....	75
4.7.8. Intra-personal intentional identity .....	77
4.7.9. Strange triangulation conditions, the reporter and the evaluator.....	77
4.7.10. Sortal relativity .....	78
4.7.11. Indexing to an agent at a time.....	79
4.7.12. Toy examples and the overall strategy .....	80
4.8. Objections and replies .....	80

4.8.1. The ‘too much discretion’ objection.....	80
4.8.2. The desire for an objective standard of subject matter individuation.....	82
4.8.3. Implicit beliefs and cognitive sophistication .....	83
4.9. Conclusion.....	85
Chapter 5: Explanatory Belief or Shared Belief? We Have to Choose .....	86
5.1. Two roles for belief .....	86
5.2. Conflation.....	86
5.3. The next step .....	90
5.4. A cautionary tale .....	93
5.5. The cost of explaining behaviour.....	97
5.6. Options .....	99
5.6.1. Settling.....	99
5.6.2. Pluralism.....	100
5.6.3. Off the well beaten track .....	104
5.7. Conclusion.....	105
Chapter 6: Putting the Triangulation Theory to Work.....	107
6.1. An unorthodox approach.....	107
6.2. The descriptivist theory of the content of attitudes .....	107
6.3. Content and subject matter.....	110
6.4. Regimenting language.....	113
6.5. Keeping the putative target stable across time and across agents .....	113
6.6. Disagreement and mind-changing.....	114
6.7. Agreement and the stability of complete attitudes .....	115
6.8. Communication .....	119
6.9. Other features of my approach.....	119
6.10. Uniformity again .....	122
6.11. Traditional arguments against the description theory of names.....	123
6.11.1. The argument from ignorance and error.....	124
6.11.2. The epistemic argument.....	127
6.11.3. The modal argument.....	127
6.11.4. The ‘passing the buck’ argument.....	128

6.12. Conclusion.....	132
Chapter 7: A Solution to Kripke’s Puzzle About Belief.....	133
7.1. Is puzzling Pierre inconsistent?.....	133
7.2. Yes and No.....	134
7.3. Unsatisfactory accounts.....	135
7.4. My proposal.....	140
7.4.1. Beliefs about identity.....	141
7.4.2. A theory of intentional identity .....	141
7.5. My solution .....	142
7.6. Conclusion.....	144
Chapter 8: Summing Up and Directions for Future Research .....	146
8.1. The first goal .....	146
8.2. The second goal.....	147
8.3. Other applications .....	149
References.....	152



‘Is it true that Anna won a new car in the lottery?’

‘Yes! But her name is Kristina not Anna, it was a bicycle not a car, it was old not new and she didn’t win it, it was stolen from her.’

## CHAPTER 1: PRELIMINARIES

### 1.1. Intentional identity

Some pairs of intentional attitudes (beliefs, desires, fears, etc.) do not have a common focus. For example, my belief that London is pretty and my belief that my mother is happy are not directed at the same thing in any sense. In some cases, intentional attitudes have a common focus and there is an object at that focus. For example, my belief that London is pretty and my belief that London is populous both have London at their focus.

Geach (1967) points out that there is an intermediate sort of case. There are pairs of attitudes that have a common focus in spite of there being, apparently, no object at that focus. Different beliefs can be about the same witch, even if there are no witches. Different astronomers had beliefs about Vulcan and their beliefs had a common focus, at least in some sense. Geach introduced the label ‘intentional identity’ to pick out the relation that holds between beliefs in these intermediate cases; he says that ‘[w]e have intentional identity when a number of people, or one person on different occasions, have attitudes with a common focus, whether or not there actually is something at that focus’ (Geach 1967, 627).

Let me illustrate. Consider scenario (a) and sentence (1), variants on a case and a sentence originally introduced by Geach.

- (a) Hob and Nob live in the same village and read the same newspaper. The newspaper reports that a witch has been terrorising the village. Hob believes that a witch (the one mentioned in the newspaper) has blighted Bob’s mare and Nob believes that the witch mentioned in the newspaper killed Cob’s sow. Witches do not exist.

(1) may be true in scenario (a).

- (1) Hob thinks a witch has blighted Bob's mare, and Nob believes she (the same witch) killed Cob's sow.

In scenario (a), Hob’s belief and Nob’s belief are what I will call *empty beliefs*, in that they appear to be directed at something that does not exist (to say a belief is empty, in this sense, is *not* to claim

that it lacks content or truth conditions). Yet (1) may be true in scenario (a). The truth of (1) requires that Hob's belief and Nob's belief have a common focus; they need to be directed at the same witch, in some sense. In scenario (a), Nob's belief *is*, in some sense, directed at the same thing as Hob's belief is, in spite of there apparently being no object at which both beliefs are directed.

The attitudes in question have more than one target. Hob's belief is directed at a witch *and* Bob's mare, and Nob's belief is targeted at a witch and Cob's sow. This is an unfortunate feature of Geach's examples that I will harmlessly sidestep for the time being. When I talk of *the* target of the beliefs I should be understood as meaning the *relevant* target. In scenarios like (a) the relevant target is the witch. Good theories of intentional identity will account for the fact that attitudes are often directed at more than one target. The theory of intentional identity I present in chapter four will be able to account for this, see section 4.7.1.

When I talk about the *target* of an apparently empty attitude, I should not be taken as necessarily committing myself to some kind of non-existent, mythical or possible object. As we will see in chapter two, some do take this view of apparently empty attitudes. The way I will use the term, an attitude's having a target might merely require that the holder of the attitude *thinks* that there is an object their attitude is about. Thinking that there is an evil witch is one thing; there being a witch that you believe is evil is quite another. I will often use the phrase 'putative target' to mean the target of an attitude in this non-committal sense; adding the word 'putative' is to remind the reader that it is this sense of the word 'target' that is intended. Take any quantification over targets that occurs in the following discussion with a grain of salt. In section 4.7.1, I will explicitly cash out the notion of a 'putative target' that I am employing.

In the above case the relevant beliefs are aimed at a particular object. Attitudes that are directed at kinds, collections of objects or properties can also have a common focus in this way. For instance, two scientists might have beliefs concerning phlogiston, even if there is no such stuff. The phenomenon of having a common focus does not require that the attitudes in question be directed at particular objects.

Call the relation of having a common focus, whether or not there is an object at that focus, the *g-relation* (in honour of Geach). Let us say there is intentional identity when intentional attitudes

(beliefs, desires, fears, etc.) are g-related. Theories of intentional identity are accounts of when and why attitudes are g-related.

The label ‘intentional identity’ is somewhat misleading. It suggests an identity relation standing between intentional objects. However, as we will see in chapters two and three, this is only one way of understanding the g-relation.

## 1.2. Why we should care

There are a number of reasons to want an adequate theory of intentional identity. Having a particular putative target as a common focus is just one way for attitudes to have a common focus. Sometimes attitudes are about the same property. In a first order ethics debate, for instance, the subject matter of the relevant attitudes might be some moral property. The subject matter of attitudes can also be classes of objects. I can have a belief about cats in general that is not a belief about a particular cat and attitudes about classes of things can have a common focus in virtue of being directed at the same class of things, in some sense. Theories of intentional identity are theories of subject matter individuation, that is, theories about when and why attitudes have subject matter in common, so the correct theory of intentional identity will be important in debates in which the notion of subject matter plays a central role.

Questions of subject matter individuation are important for debates concerning agreement and disagreement. Suppose I believe that Jill is ill and you believe that Jill is not ill. In virtue of what do our beliefs conflict? A natural thought is that these beliefs conflict because they have a common focus, Jill, and they ascribe conflicting properties. Similarly with agreement, if we both believe that Jill is ill, in what sense do we agree about Jill? Once again, a natural thought is that we agree because our beliefs have a common focus, Jill, and they ascribe the same property. Clearly, the question of which attitudes have a common focus will be relevant to questions of when agents agree and disagree. Questions concerning when and about what agents are disagreeing come up in all sorts of debates in philosophy: metaphysics, philosophy of language, meta-ethics, and philosophy of mind, to name a few. A general theory of intentional identity would thereby yield insight into areas of these debates.

In disputes, philosophers often claim that this or that argumentative move *changes the subject* and

that this has some import when evaluating the argumentative move in question. In meta-ethics, for instance, it is often contended that if someone give up the claim that the target of moral discourse has certain apparently crucial properties (for example, objectivity), they are not talking about morality any more. They are supposed to have *changed the subject* from morality to something else.<sup>1</sup> Having an adequate theory of intentional identity will allow us to regiment talk of subject matter and of changing the subject, as well as help us to evaluate particular ‘changing the subject’ arguments.

Relatedly, subject matter individuation is often taken to be relevant to the issue of distinguishing genuine disagreement from verbal disagreement. The difference between these kinds of disagreement is, it seems, a difference in *subject matter*. The subject matter of a genuine disagreement is the target phenomenon itself, whereas the subject matter of a verbal disagreement is either the words or concepts used to formulate theories of the target phenomenon. If this is right, issues of subject matter individuation will be important when answering questions about whether a given disagreement is verbal, genuine or, perhaps, both.<sup>2</sup>

When do agents successfully communicate about a target entity? A natural thought is that successful communication requires that the relevant thoughts have a common subject matter. It is possible for agents to communicate about things even though the relevant thought and talk is empty. For example, when planning their expedition, explorers communicated about the fountain of youth, sharing their views about where it was to be found. If communication of this kind requires that the relevant thought and talk has a common focus, then having an adequate theory of intentional identity will allow us to better understand this kind of communication and, I think, communication in general.

Finally, and perhaps most importantly for the project of this thesis, answering questions about how we ought to understand intentional identity will also yield insight into how we ought to approach

---

<sup>1</sup> See, for instance, Joyce (2008) and part 6 of Parfit (2011).

<sup>2</sup> Using the labels ‘genuine’ and ‘verbal’ might suggest that verbal disagreements are not serious or interesting. This is not the case. Some verbal disagreements are extremely important. For instance, a debate about what the word ‘torture’ should mean in international law is extremely interesting and significant despite being verbal.

more general questions of intentionality. Intentional identity is closely tied to general issues concerning *aboutness* or *intentionality*. Developing a theory of intentional identity will, I think, lead to very interesting conclusions about the content and subject matter of intentional attitudes. Cases like Geach's have two properties that make them tricky to handle. The relevant intentional attitudes involved are empty *and* they have a common focus. This means that an adequate account of cases like Geach's require us to have both an adequate theory of empty attitudes *and* an adequate theory of intentional identity.

Cases involving empty intentional attitudes have been a major stumbling block for some theories of the content of attitudes (e.g. naïve Russellian theories defended Salmon (1986, 1989, 2006), Braun (1998, 2001b, a), Soames (1987) and Tye (1978) among others) and one of the main reasons in favour of others (e.g. the descriptivist theory of the content of attitudes defended by Russell (1905, 1910-1911) and Quine (1948)).

One of the goals of this thesis is to defend a certain individualistic approach to the content of attitudes. Individualism about the content of intentional attitudes is, roughly speaking, the doctrine that what attitudes an agent has and the content of those attitudes supervene on what is going on under their skin. The main kind of rival to the approach I favour involve appeals to externalist constraints on the content of beliefs. Many believe that the world helps intentional attitudes get their content, in some sense. Some think intentional attitudes get to concern objects in the world via causal chains that stand between some objects in the world and the attitude. This sort of view flows nicely from the work of Donnellan (1970, 355-356), Putnam (1975), Kripke (1980, 96-97) and others. Others claim that certain worldly entities are just more eligible to be involved in the content of intentional attitudes than others and that this fact does not depend in any way on the agents with the attitudes or their relationship to these objects. A well-known example of a theory with this general structure is discussed by Lewis (1983, 370-377, 1984, 227) and similar theories are defended by Sider (2009) and Weatherson (2003). These are examples of views that involve appeals to worldly facts that make a difference to the content of intentional attitudes or to what attitudes are about.

In cases like (a) the thing that the relevant attitudes are apparently about is missing. Even if there are some kinds of merely possible, mythical or non-existent objects that the attitudes are about, these objects do not appear to have a place in the world to which we might appeal when claiming

that a given apparently empty belief is about this or that object. For example, the objects do not make a causal difference to concrete objects at the actual world. The usual externalist constraints do not appear to be in play. *And yet*, the attitudes have content and appear to have a common focus.

My argumentative strategy will go something like this: we need a theory that accounts for the content of attitudes in cases like (a) and accounts for how apparently empty attitudes can have a common focus. Whatever explanatory resources this theory requires, it will not involve the usual externalist resources (causal relations between the target object and the holder of the attitude, or eligibility constraints) since they are missing in cases like (a). I will then argue that we ought to apply the theory developed for cases like (a) (that does not involve appealing to externalist constraints) to *all* cases.

There are at least two reasons why we ought to apply whatever theory of attitude content and intentional identity adequately accounts for cases like (a) to all cases. The first is the pressure for uniformity. All else being equal, we ought to prefer theories of attitude contents and intentional identity that treat empty and non-empty intentional attitudes in the same way. If we applied the theory that works in cases like (a) to all cases we could maintain a uniform treatment of intentional identity and the content of attitudes.

The second is that appealing to external constraints on attitude contents is costly and if we can avoid those costs we should. I will discuss these costs at length in chapter five. If I am right, there is pressure to apply an account that *does not* appeal to these constraints, since, in doing so, we can avoid the costs of appealing to external constraints altogether. In this way, an adequate theory of intentional identity could come to form the centrepiece of a broader approach to questions concerning intentional attitudes.

Finally, there appears to be close to a dialectical stand-off in some debates concerning attitudes in the easier cases in which there is an obviously existing object that is uncontroversially connected, in some sense, to both attitudes. In chapter seven I will discuss some such cases. Considering what we should say about the harder cases in which the relevant attitudes are empty will, I think, give us guidance about what to say about the easier cases and do something to break the standoff.

### 1.3. What we are looking for in a theory of intentional identity

What should we look for when evaluating theories of intentional identity? One obvious thing is extensional adequacy. We want a theory of intentional identity that systematically delivers, as far as is possible, the correct verdicts concerning intentional identity in particular cases. In turn, we take our guidance about which verdicts are correct from our pre-theoretic judgements about cases.

Another desirable feature of a theory of intentional identity is that it provides a uniform treatment of cases involving empty and non-empty attitudes. This lines up with Geach's gloss on intentional identity, quoted above. He describes intentional identity as the relation of having a common focus *whether or not* there is an object at that focus, this suggests that a good theory of the relation would treat intentional identity the same in both cases. Uniformity considerations can cut in either direction. In chapters three and four there will be some discussion concerning which theories do and do not deliver this sort of uniformity.

If what I said in the previous section is correct, then intentional identity is closely linked to a wide range of phenomena. For this reason, it would be preferable if our theory of intentional identity hooked up with those phenomena in the right way. For instance, if intentional identity is relevant to issues concerning disagreement, it would be preferable if our theory of intentional identity could also yield constraints on a theory of disagreement. It is no good to come up with a theory of intentional identity that, if it were correct, severs the link between intentional identity and phenomena to which it is, in fact, closely related.

Another move that we ought to try and avoid is a kind of primitivism or epistemicism concerning the relation of intentional identity. That is to say, we ought to stick to *explanatory* theories of intentional identity, as far as possible. We might be tempted, to claim when confronted by the challenges that face particular theories of intentional identity, that the relation of aboutness and the relation of having a common focus are, somehow, primitive and unanalysable. If this claim is true we should stop trying to develop explanatory theories of the relation. At the very least, we would be forced to approach the question of intentional identity very differently.

Another reason we might be tempted to give up on finding an explanatory theory of intentional identity is if we accepted a kind of epistemicism about intentional identity. I have in mind the kind of epistemicism defended by Sorenson (1988). On this sort of story, there is some fact of the matter



concerning when attitudes have a common focus but it is, somehow, inaccessible.<sup>3</sup> Again, if this sort of view is right we may as well stop looking for an explanatory theory of intentional identity. I suggest that we not give up the hunt for an explanatory theory of intentional identity quite so hastily, although I admit that if, after a thorough investigation into intentional identity, we find that the phenomenon obstinately resists analysis, then the primitivist and epistemicist options will perhaps start looking more attractive.

Many other good-making features for theories of intentional identity will come up as we progress but the above desiderata will be enough to get us started in evaluating theories of intentional identity.

#### 1.4. What the project is *not*

Theories of intentional identity are accounts of when and why attitudes are g-related. The project of finding an adequate theory of intentional identity is *not* the project of providing a systematic account of how intentional identity is ascribed in language. In the literature on intentional identity much attention<sup>4</sup> has been paid to the grammatical and syntactic features of sentences ascribing intentional identity. (1), for instance, contains an anaphoric pronoun ('she'), a feature it has in common with many sentences that ascribe intentional identity. If one takes a brief survey of the literature one may be forgiven for thinking that those working on intentional identity consider the question of the semantics of anaphoric pronouns to be inextricably linked to the question of when and why there is intentional identity. Indeed this comes across strongly from the way that Geach originally poses the puzzle, in terms of a sentence that needs to be adequately analysed. But this is not the case. We need only consider that some sentences ascribe intentional identity without containing anaphora. For instance, consider (2)

(2) Sal and Val fear the same werewolf.

(2) may be true even when there are no werewolves.

It is one thing to understand what it is for attitudes to have a common focus, but it is another to

---

<sup>3</sup> Thanks to Graham Priest for bringing this option to my attention.

<sup>4</sup> See, for instance, King (1993), Cumming (2014) and Lanier (2014).

understand precisely how this relation is signalled in language. The second issue is, no doubt, interesting and important but it is not the focus of this thesis.

However, the issue of how intentional identity is signalled in language will come up later. There are two reasons it will come up. The first reason is that many accounts of intentional identity are both accounts of the relation *and* accounts of the semantics of the sentences used to ascribe the relation, such as King's (1993, 65) view, Manning's (2015) view and some of the views summarily rejected by Geach (1967, 630). So in evaluating these theories we will have to engage with them on their level.

The second reason is more interesting. Sometimes facts about the semantics of intentional identity sentences make salient certain crucial features of intentional identity. We will see an example of this in chapters three and four, when we consider a case presented by Edelberg (1986, 1992). A theory of the semantics of intentional identity sentences may put constraints on theories of intentional identity but it does not *constitute* a theory of intentional identity.

## 1.5. The plan

This thesis has two main goals. The first is to develop a theory of *intentional identity*. Chapters one through four inclusive will be devoted to evaluating the theories of intentional identity to be found in the literature, and developing a new theory of intentional identity. In chapters two and three I will argue that the extant theories of intentional identity in the literature are, in some way or other, inadequate. In chapter four I will present a new theory of intentional identity, *the triangulation theory*, and argue that it has many significant advantages over its principal rivals.

The second goal is to use the triangulation theoretic tools I developed in chapter four to shed light on other debates concerning intentional attitudes. In chapter five I argue that there is a tension between two roles the notion of belief is supposed to play. On the one hand, beliefs are supposed to help us explain the behavior of the believer, but on the other hand, beliefs are supposed to be the sort of things that are often shared by agents and by single agents across time. As it turns out, accounts of belief that allow belief to play one of these roles make them unable to play the other. In chapter six, I use triangulation theoretic tools to resolve this tension in a novel way. Triangulation theoretic tools can help us reconcile the claim that attitudes are often stable across agents and

across time with the claim that attitudes are not often shared between agents and across time. In chapter six I show how these tools provide answers to answer a whole swath of objections to descriptivist theories of the content of intentional attitudes. In chapter seven I argue that the triangulation theory also yields an attractive solution to Kripke's puzzle about beliefs. In Chapter eight I will sum up the discussion and discuss some ways in which triangulation theoretic tools can shed light on other philosophical debates outside of the philosophy of mind and intentionality. I will also mention avenues for future research suggested by the proceeding discussion.

## CHAPTER 2: WHICH WITCH IS WHICH? EXOTIC OBJECTS AND INTENTIONAL IDENTITY

### 2.1. Intentional identity and transparency

My belief that London is populous and my belief that London is pretty have a common focus. It is natural when confronted with ordinary cases like these, to suggest that, in general, attitudes have a common focus *because* there is an object that they are both about.

This suggestion is put under pressure by cases in which attitudes appear to have a common focus in spite of there being, apparently, no object at that focus. Scenario (a) is one such case, repeated here for ease of reference.

- (a) Hob and Nob live in the same village and read the same newspaper. The newspaper reports that a witch has been terrorising the village. Hob believes that a witch (the one mentioned in the newspaper) has blighted Bob's mare and Nob believes that the witch mentioned in the newspaper killed Cob's sow. Witches do not exist.

(1) may be true in scenario (a).

- (1) Hob thinks a witch has blighted Bob's mare, and Nob believes she (the same witch) killed Cob's sow.

Hob's belief and Nob's belief appear to be directed at something that does not exist. Yet in scenario (a), Nob's belief *is* directed, in some sense, at the same thing as Hob's belief is, in spite of there apparently being no object that both beliefs are about.<sup>5</sup>

Cases like this might lead us to contend, with Geach (1967, 627), Dennett (1968), Perry (2001, 14,

---

<sup>5</sup> I will focus on attitudes with particular objects as targets, since considering these cases will be sufficient for me to put forward my arguments. However, a good theory of intentional identity will also be adaptable to cases in which beliefs are directed at classes of objects, or at kinds.

147-156), Glick (2012), Sainsbury (2005, ch. 7, 2010), and van Rooy (2000, 170-178) that attitudes can have a common focus even when there is no object at that focus; the presence of an object that attitudes are about is *not* required for their having a common focus.

Others such as Salmon (2005, 105–108), Parsons (1974, 577–578), and Saarinen (1978, 207–210), whom I call *transparent theorists*, deny that cases like (a) mean that attitudes can have a common focus without being about the same object. Instead, they claim that apparently empty attitudes are, appearances notwithstanding, about objects of some kind. I will often drop the ‘apparently’ from ‘apparently empty’ for the remainder of the thesis. When engaging with transparent theorists it is important not to take the use of the word ‘empty’ too seriously. Different transparent theorists say different things about the kind of objects that apparently empty attitudes are about but the central strategy is the same: claim that empty intentional attitudes are about exotic objects of some sort (usually merely possible, abstract or non-existent objects) and claim that having a common focus depends on the attitudes having a common object, even in the face of cases like (a). According to transparent theories of intentional identity, non-empty attitudes get to have a common focus in virtue of their being about the same exotic object. Another way of understanding the spirit of the transparent view is that the subject matter of an attitude is just what that attitude is *about*, and this is all there is to subject matter individuation.

There are several benefits associated with understanding intentional identity in this way. The transparent account fits well with ordinary talk concerning empty beliefs. Consider the case in which astronomers had beliefs concerning Vulcan. We naturally describe these beliefs as being *about Vulcan* and it is natural to describe several of those beliefs as being *about the same thing*, namely Vulcan. The transparent theorist can take these locutions at face-value.

Transparent theories deliver a uniform treatment of intentional identity across empty and non-empty attitudes. For the transparent theorist, empty beliefs are directed at objects in much the same way non-empty beliefs are. Having a common focus is, for the transparent theorist, always a matter of the attitudes being directed at the same thing, whether or not the attitudes in question are empty. All else being equal, we ought to prefer theories that treat empty and non-empty attitudes uniformly. This is a point in favour of adopting a transparent approach to intentional identity.

The transparent theorist is able to account for communication and co-ordination of action, in a

natural and elegant way, in cases in which the relevant attitudes are empty. Agents can communicate even when their relevant attitudes are empty. Astronomers successfully communicated about Vulcan and we ought to be able to make sense of this communication. What is more, agents can co-ordinate their actions even when the object at which their actions are directed isn't there; bands of explorers undertook significant joint projects while searching for the fountain of youth, even though there is no such thing. Transparent theorists can handle these phenomena in a straightforward way; agents are able to communicate about an object and direct their actions toward that object in virtue of their relevant attitudes being *about* the very same object.

Transparent theories allows us to understand, in a straightforward way, how agents could genuinely disagree about a thing even when the relevant attitudes are empty. For example, explorers were able to disagree about where the fountain of youth was even though their attitudes concerning that object were empty. If we adopt a transparent account of intentional identity we can understand this disagreement in a straightforward way; the subjects disagree in cases like this because there is an object about which they have conflicting attitudes.

In this chapter I argue that, despite these attractions, a transparent approach to intentional identity ought to be rejected. We ought not to posit exotic objects in the service of explaining intentional identity. It may be that we ought to be committed to these exotic objects for some *other* reason, but positing them does not yield any distinctive resources for understanding intentional identity. The reason, in a nutshell, is that there are too many exotic objects and not enough facts to which we can appeal when determining which exotic objects are assigned to which attitudes. When the objects in question are actual and concrete we can appeal to the object's place in the world (e.g. the causal relations it stands in) when explaining why a given belief is about that object rather than some other. But when the objects in question are *exotic* there are no such resources to which to appeal. This has important implications for the prospects of transparent approaches to intentional identity.

I will focus on a particular type of intentional attitude: *belief*. Considering beliefs will be sufficient to bring out the problem, but the arguments could just as easily be made with cases involving other types of intentional attitudes.

The plan for the rest of this chapter is as follows: first I say more about the sort of exotic objects

to which different transparent theorists appeal. I then discuss what is required if a transparent theory is to be an adequate theory of intentional identity. Next, I argue that transparent theorists run into trouble when they try to deliver the correct verdicts about certain sorts of cases. I then discuss and evaluate some responses to the problem available to transparent theorists. Finally, I draw out some lessons that ought to be taken from the discussion in this chapter.

## 2.2. Exotic objects

As already mentioned, different transparent theorists say different things about the kind of objects that empty intentional attitudes are about. One sort of transparent theory, a version of which is defended by Salmon (1998, 2005), involves an appeal to abstract mythical objects. If this is right, we can claim that two empty beliefs have a common focus just in case they are about the same mythical object. According to this sort of view, the truth of (1) does not require the existence of a concrete, flesh and blood witch, it merely requires the existence of an abstract mythical witch.

At this point, we might distinguish between *fictions* which do not pretend to be anything but fiction and *myths* which are mistakenly believed to be correct descriptions of the world. If the mythical object account is only supposed to handle cases involving fictions in this sense, the account does not constitute a complete account of intentional identity. There will be cases of intentional identity in which the relevant agents sincerely believe in the existence of the objects in question. I will, therefore, only be concerned with the more general type of account, the type that applies both to objects of fiction and objects of myth, for the purposes of this chapter.

Among those who countenance mythical objects we can distinguish Platonists from creationists. According to Platonists, abstract objects are eternal; subjects simply select from the set of abstract objects when they have empty beliefs and the relevant beliefs are about whatever object is selected. Creationists, in contrast, say that some abstract objects do not exist at all times but are created when certain conditions are met.<sup>6</sup> For instance, Vulcan, the mythical object, was literally created by le Verrier when he attempted to explain certain facts about Mercury's orbit by positing the

---

<sup>6</sup> Some notable creationists are Thomasson (1996, 300–307, 1999, 5–14), Zvolenszky (2012, 2014), Salmon (1998, esp 293–296), and Kripke (2013, 71–78).

existence of a planet he called ‘Vulcan’. This distinction will be important below.

Another kind of transparent theory involves appealing to non-existent objects. For instance, Parsons (1974, 577–578) and Saarinen (1978, 207–210) claim that apparently empty beliefs are about non-existent objects, and that empty beliefs have a common focus just in case there is a non-existent object that they are both about. If this is right, the truth of (1) need not require the *existence* of an object that Hob’s belief and Nob’s belief are about, since beliefs can be about non-existent object.

Transparent theorists might instead appeal to merely possible objects. In line with this proposal, we would claim that empty intentional states are about possible objects and that intentional identity is a matter of beliefs being about the same object, actual or possible. We could then understand intentional identity between empty beliefs as depending on attitudes being about the same merely possible object. The truth of (1) would then not require the existence of witches in the actual world, rather it merely requires the *possibility* of a witch.

Before I proceed, I wish to mention a theory of intentional identity only to set it aside. I have in mind what we might call the ‘scapegoat’ theory, which is a kind of transparent theory. According to the scapegoat theory, there is intentional identity just in case there is an object that both attitudes are about but the objects in question are *always* actual, concrete, existent objects. On this kind of view we might say that (1) is true in scenarios like (a), just in case there was some person or object that Hob and Nob both suspect of witchcraft. Slater (2000) defends a theory of this kind. The reason I will set this sort of account aside is that it has bad implications and delivers the wrong verdicts about intentional identity even in simple cases like (a). In particular, this account is inadequate because (1) may be true in scenarios like (a) in which Hob and Nob have no particular suspicions about who is practicing witchcraft.

### 2.3. Assignment

For transparent theorists, intentional identity depends on the presence of an object that both attitudes are about. However, if we want an adequate account of intentional identity it is not enough to claim that when the beliefs are about the same object there is intentional identity. We need a story about when and why attitudes get to be about the same object. Without an account of this



sort transparent theories do not yield verdicts about intentional identity in particular cases. Why, for example, are Hob's belief and Nob's belief about the same object in scenario (a)? We are told that if (1) is true they are about the same object, but we are not told *when* and *why* they are about the same object.

There is another relevant question that, if it were answered, would provide an answer to the question of when attitudes are about the same object: why are particular attitudes about the objects they are about? If we determined which objects attitudes are about, we would *also* determine which attitudes are about the same object.

Call an account of why attitudes are about the objects they are about an *assignment principle*. An assignment principle is an account of how objects are *assigned* to attitudes. What sort of assignment principle a transparent theorist adopts will determine which verdicts, concerning the presence or absence of intentional identity, their theory will yield in particular cases; there will be intentional identity when attitudes are assigned the very same object. Transparent theorists are, for the most part, not explicit about the sort of assignment principle they adopt.

#### 2.4. The hard cases

The argument I present below concerns what I will call *the hard cases*. The hard cases are the cases, like scenario (a) presented above, in which empty attitudes have a common focus. I will argue that transparent theorists are unable to satisfactorily account for intentional identity in the hard cases.

Now, it may be that they will be unable to provide an assignment principle that delivers the correct verdicts about intentional identity even in non-empty cases. The issue of how non-empty attitudes get to be about the objects they are about is, after all, notoriously vexed. However this may be, if I am right, the transparent theorist's resources will not yield an assignment principle that delivers the correct verdicts in certain hard cases. Since what is distinctive about transparent theories is their treatment of the hard cases, their getting these cases wrong is enough to undermine the transparent approach.

To make my case I need to make two observations about the hard cases and one suggestion about

what might constrain our assignment of objects to empty beliefs. With this background in place I will be able to present a problem for transparent theories.

#### 2.4.1. Abundance

The first observation is that, if we are committed to such things, there is an abundance of non-existent objects, possible objects and, at least for the Platonist, mythical objects; in the hard cases there are a huge number of exotic objects to choose from when assigning objects to attitudes. Those who appeal to non-existent objects, for instance, are fond of saying that for (almost)<sup>7</sup> any set of properties, there is an object, existent or non-existent, which has those properties. The set of possible objects is also abundant; for every way a thing could be there is a possible object that is that way. For the Platonist about mythical objects there is also an abundance of mythical objects. All the mythical objects we could ever dream up, and more, exist as abstracta. Creationists about mythical objects are in a position to claim that the set of mythical objects is, for them, less abundant as only those objects that have been created exist. I will set aside the moves that are uniquely available to the creationists until section 2.6.2.

If we are to understand empty beliefs as being about objects, an abundance of objects is necessary if we are to capture the content of *all* empty beliefs that appear to be directed at objects. For any such empty belief there needs to be at least one corresponding object. If a transparent theorist takes some non-existent, merely possible or mythical objects out of the picture so that they cannot, even in principle, be assigned to empty beliefs, they will lose the ability to capture the content of some empty beliefs, in particular those empty beliefs that are best interpreted as being about those objects. To put it another way, we should never run out of objects when assigning objects to empty beliefs. There should be no point at which our metaphysics of objects comes in and dictates that we cannot have this or that object-directed empty belief. If a transparent theorist significantly limits the relevant set of objects, she would limit the expressive power of her theory.

#### 2.4.2. Poverty

---

<sup>7</sup> The constraints Meinongians put on this claim need not concern us here.

The second observation is that in the hard cases there are very few resources to which the transparent theorist might appeal when assigning exotic objects to attitudes. There are very few resources available to explain why one exotic object rather than another is assigned to a given attitude. One way to bring out this poverty in the hard cases is by considering the resources that *are* apparently available when assigning actual, concrete, existing objects to non-empty beliefs.

When deciding which objects to assign to a non-empty belief, we can often appeal to the object's place in the world and its relationship to the believer. For instance, why are some of my beliefs about London rather than Paris? A natural thought is that London, but not Paris, stands in the right sort of causal relation with my London beliefs, and if we make our assignment principle sensitive to that fact, then we can explain why my belief is assigned one city rather than the other. London and Paris are causally related to me in different ways, and this fact might explain why my London beliefs are assigned the object London rather than Paris. This sort of view flows nicely from the work of Donnellan (1970, 355-356), Putnam (1975), Kripke (1980, 96-97), and others.

Another kind of resource that is sometimes appealed to in non-empty cases are eligibility facts. According to this idea, some objects are just more eligible than others to be what attitudes are about. Some contend that certain actual, concrete and existent objects are more natural than others, in that they have boundaries that carve nature relatively close to its joints and we can appeal to these sorts of facts about naturalness when assigning objects to beliefs. If this is right, a given belief can be about an object *x* and not about an object *y* in virtue of the fact that *x* is more eligible than *y*. A well-known example of a theory with this general structure is discussed by Lewis (1983, 370–377, 1984, 227) and somewhat similar theories are defended by Sider (2009) and Weatherson (2003).

But when deciding which exotic object to assign to an empty belief we cannot, it seems, appeal to these sorts of causal facts and eligibility facts. Exotic objects do not have a unique place in the world in the same way that London and Paris do. The objects in question appear not to have a place in the causal order of things; non-existent objects do not causally interact with existent objects. Merely possible objects such as possible witches or possible planets appear not to make a causal difference to the actual world. For the Platonist about mythical objects at least, mythical objects do not causally interact with concrete objects like Hob and Nob. For both the creationist

and the Platonist, abstracta do not make a causal difference to concrete objects or events. For these reasons, we cannot, it seems, appeal to causal constraints when we are assigning one possible, non-existent or mythical object to an empty belief rather than another.

It is also unclear whether there are the right kind of relations of comparative eligibility standing between these exotic objects. It is intuitive and plausible that an object like Barack Obama has boundaries that match objective joints in nature better than an object that has Barack Obama and Jupiter as parts. Maybe planet-people are less eligible than planets or people. The trouble is that there do not seem to be the appropriate objective joints in the space of exotic objects. After all, not just any eligibility constraint will help the transparent theorist answer the assignment challenge. An eligibility constraint, if it is going to help, must distinguish different objects and not just different *kinds* of objects. We are interested in which witch is which, not in the distinction between witches and some other kind of exotic object. It may be, for instance, that merely possible, abstract or non-existent witches are more eligible to be what beliefs are about than merely possible, abstract or non-existent shwitches (where ‘shwitch’ is a relatively gerrymandered kind), but this kind of relative eligibility will not help much with the assignment challenge. We are looking for an explanation of why a *particular object* is assigned to an attitude, not an explanation of why a certain *kind* of object is assigned to the belief; we want a belief about Vulcan, not merely a belief about planets.

According to another conception of eligibility, some entities are relatively eligible in virtue of their having a relatively simple definition or description in terms of certain special fundamental properties. The non-existent, possible or mythical witches do not appear to be related to fundamental properties of this sort, at least not at the actual world. It is not obvious, for instance, that these exotic objects stand in the same kind of relation to fundamental properties as the tree outside my office does. It is even less clear that any *one* witch is more intimately related to the fundamental properties than another witch is.

There may be good reasons to be suspicious of appeals to eligibility facts, even in non-empty cases. However, whatever one’s attitude to eligibility constraints in non-empty cases, one should at least be *more* suspicious of views that appeal to eligibility facts among non-existent, merely possible or mythical objects.

In the non-empty cases it is at least open to us to appeal to causal constraints or eligibility constraints on assignment, though appeals to these constraints are by no means straightforward or uncontroversial.<sup>8</sup> The hard cases are *impoverished* in that these resources appear to be absent; the objects in question do not have a distinctive place in the world to which we can appeal when assigning them to beliefs.

### 2.4.3. The natural suggestion

What facts *can* the transparent theorist appeal to when assigning exotic objects to empty attitudes? A natural suggestion is that how the believer takes the target of their belief to be puts some constraints on which objects we ought to assign to that belief. For instance, why are beliefs concerning Vulcan about a non-existent, merely possible or mythical planet rather than a non-existent, merely possible or mythical elephant? A natural suggestion is that the believers take the target of their beliefs to be a planet and not an elephant. Le Verrier conceived of a planet, not an elephant, and those who have beliefs concerning Vulcan tend to believe that the target of their belief is a planet. Call the way that the believer takes the target of their belief to be, the properties they believe it to have, the believer's *characterisation* of the target.

*The natural suggestion* as I will call it, says that the believer's characterisation of the target of their belief constrains which non-existent, merely possible or mythical objects ought to be assigned to that empty belief, at least in the hard cases. We may appeal to the extent to which different objects *fit* the believer's characterisation of the target when assigning exotic objects to beliefs.

There is a puzzle about how the objects of empty beliefs are supposed to have the properties they are characterised as having. In what sense is Vulcan a planet? In what sense did Sherlock Holmes live in London? I will put these issues to one side. Even if the objects do not, strictly speaking, have the properties they are characterised as having, there will presumably be some principled way

---

<sup>8</sup> See for instance Kroon's (1987) critique of causal theories of reference. The message of Kroon's excellent paper is that until those who appeal to causal constraints on reference specify *which* causal chains make for assignment in particular cases, the causal theory ought not to be considered adequate. This is a challenge that causal theorists are, as far as I know, yet to adequately face.

of talking about members of the relevant set of objects *fitting* characterisations.

The abundance observation, the poverty observation, and the natural suggestion are somewhat compelling and I will work in accordance with them for the time being. I will discuss some ways they might be resisted in section 2.6.1.

## 2.5. The problem

With the abundance observation, the poverty observation, and the natural suggestion in place, I can make trouble for the transparent theorist. First consider scenario (b).

- (b) Hob and Nob live in the same village and read the same newspaper. The newspaper reports that a witch has been terrorising the village. Hob believes that the witch mentioned in the newspaper blighted Bob's mare. Nob thinks the witch mentioned in the newspaper did not blight Bob's mare but did kill Cob's sow. Witches do not exist.

(1) is true at scenario (b); Hob's belief and Nob's belief have a common focus. For the transparent theorist to deliver this verdict about scenario (b) they must provide an assignment principle that assigns the same object to Hob's belief and Nob's belief.

What kind of assignment principle could deliver this verdict? In line with the natural suggestion we might say that the extent to which an object fits the agents' respective characterisations of the target can play some role in determining which object the belief in question is about. Hob and Nob characterise the target of their beliefs in different ways. Hob thinks the target of Hob's belief blighted Bob's mare whereas Nob thinks the target of Nob's belief did not blight Bob's mare. According to abundance, there are two distinct witches, one that fits Hob's characterisation but not Nob's characterisation and another that fits Nob's characterisation but not Hob's. These two witches will fit the respective characterisations better than any particular that witch fits both. If all we have to go on when assigning objects to Hob's belief and Nob's belief is the way they characterise the putative target *and* Hob and Nob characterise the target of their beliefs differently, then these two objects ought to be assigned to Hob's belief and Nob's belief respectively.<sup>9</sup> Thus, (1)

---

<sup>9</sup> One option at this point is to posit inconsistent objects. An inconsistent object might fit both characterizations in spite of their being inconsistent. This option is misguided. So long as we are

would come out false because the relevant beliefs would be assigned different objects. But this is the wrong result, at (b) Hob's belief and Nob's belief do have a common focus; (1) is true in that scenario. So fit constraints alone cannot yield *necessary* conditions on intentional identity.

There is another kind of case that shows that, given the poverty and abundance observations, considerations of fit cannot supply *sufficient* conditions on two attitudes being assigned the same object. Consider scenario (c).

- (c) Hob, a butcher in Arkansas, believes that a witch lurks under his bed and she has blighted the mare of his neighbour Bob. Nob, a baker in Ukraine, believes that a witch lurks under his (Nob's) bed and that she killed the sow of Nob's neighbour Cob. Hob and Nob have never met and there has been no relevant interaction between their social circles. Witches do not exist.

(1) is false in scenario (c); Hob's belief and Nob's belief do not have a common focus. So, if a transparent account is correct, there ought *not* be an object that both beliefs are about in scenario (c).

There is, if the abundance observation is correct, a non-existent, possible or platonic mythical object that fits both Nob's characterisation of the target and Hob's characterisation of the target. It lurks under both beds and maligns both animals. If all we have to go on when assigning objects to these beliefs are considerations of fit, then this merely possible, non-existent or mythical witch (the one that lurks under both beds and maligns both animals) can be assigned to both beliefs, since it fits both characterisations. Nothing in either characterisation rules this object out, and since there appear to be no other constraints on assignment that could rule this object out, we may assign that object to both beliefs. But this is the wrong result. Intuitively, Hob's belief and Nob's belief do not have a common focus in scenario (c); (1) is false at (c). What is more, cases like (c) are commonplace. There is nothing special about witches, Arkansas or Ukraine; assuming abundance,

---

generous about inconsistent objects it would threaten to make *any* two empty attitudes have a common focus, since no amount of difference in characterization could possibly rule out assigning the same object to both attitudes.

for any consistent pair of characterisations, there will be an exotic object that fits both characterisations. Cases like (c) that bring out this problem are, therefore, likely to be extremely easy to cook up.

These two sorts of cases show that, if poverty and abundance are both correct, considerations of fit cannot supply both necessary *and* sufficient conditions on intentional identity. The transparent theorist cannot come up with an assignment principle that will allow them to deliver the correct verdicts about *both* cases like (b) and cases like (c).

If we could appeal to some constraint other than fit we could avoid these hazards. For instance, if we could appeal to causal constraints we could say that one witch is causally connected with both beliefs in scenario (b). This could allow us to assign the same object to the relevant beliefs in spite of the differences in how Hob and Nob characterise the target. Or perhaps if we could appeal to eligibility constraints we might be able to rule out the witch that lurks under both beds at (c) as the object that we ought to assign to the beliefs on the grounds that she is relatively ineligible.

But if the poverty observation is correct then no such resources are available in the hard cases. Apparently only fit constraints are available and, if the above argument is right, an assignment principle based only on fit cannot deliver the correct verdicts about cases like (b) *and* cases like (c). If the presence of an object that both attitudes are about is going to play a distinctive role in explaining intentional identity, we need a way of linking the relevant exotic objects to intentional attitudes that does not just involve considerations of fit.

## 2.6. Responses

What avenues of response are available in the face of this challenge?

### 2.6.1. Resisting abundance, poverty or the natural suggestion

The most obvious response is to resist one of the ingredients of my argument: abundance, poverty or the natural suggestion. We might try denying abundance. However, as discussed above, this move comes at the cost of crucial expressive power. If we take some objects out of the picture so that they cannot, even in principle, be what empty beliefs are about, we will lose the ability to



capture the content of some empty beliefs in line with the transparent account. Unless the transparent theorist wants to sacrifice expressive power, she ought to accept the abundance observation.

Should the transparent theorist deny the natural suggestion? To begin with, the natural suggestion is extremely plausible. It seems that how the believer takes the target of their attitude to be does put some constraints on which objects those attitudes are about. However, even if the natural suggestion is rejected we are no closer to an account of how empty intentional attitudes get to be about the exotic objects they are about. I was driven to make the natural suggestion because there appeared to be no other constraints on assignment that the transparent theorist could appeal to in the hard cases. Rejecting the natural suggestion makes the challenge of providing an assignment principle even more pressing.

Another option is to resist the poverty observation. But what other constraints on assignment might a transparent theorist appeal to in the hard cases? Particular non-existent, merely possible or mythical objects do not seem to make a difference to actual, concrete events and objects, so appealing to causal relations when assigning objects to empty beliefs seems ill advised.

An eligibility constraint of the right sort, if it could be imposed, would allow a solution to the problem. Recall that the eligibility constraint, if it is going to help in cases like (b) and (c), must distinguish between particular objects, not just between kinds of objects. There does not appear to be any independently plausible way of construing eligibility such that it both applies to the relevant domain of objects and allows the appropriate distinctions between objects to be made.

Another possibility is that I am not being sufficiently imaginative. There may be some other facts to which a transparent theorist could appeal when assigning objects to empty attitudes that would allow a transparent approach to be vindicated in the face of the challenge presented above. I will leave it to transparent theorists to attempt to provide an account of a constraint on assignment that will fit the bill. I am, however, sceptical about the prospects of this approach.

### 2.6.2. Distinctive creationist options?

Are there any moves, available uniquely to creationists, which will make it easier for them to provide an assignment principle that delivers the correct verdicts about cases? Creationists are in

a position to resist the claim that the set of mythical objects is as abundant as the Platonist would have us believe. For the creationist, the only mythical objects that exist at any given time are those that are supported by fictions at that time. The creationist might hope that this relative scarcity of objects might make the assignment challenge easier to answer.

Consider scenario (d).

- (d) A child comes to believe that a monster lives under her bed and that the monster is hairy. No one in the child's community has ever told her about monsters that live under beds. In fact, there has been no prior thought or talk in the child's community about monsters that live under beds. There is no monster under her bed.

The child's belief at (d) is a prime example of an empty belief that is directed at a particular object. If the creationist's account of mythical objects is to capture the content of empty beliefs in general, it had better be that constraints on when a mythical object is created are weak enough such that the belief described in (d) has an object to be about. If the creationist denies this, she loses the ability to correctly capture the content of some beliefs, for instance the belief mentioned in (d). On the other hand, if the creationist *does* intend to capture the content of the belief in (d), she is forced to make the conditions on creation very weak; all the child did was come to believe things about what lives under her bed. Salmon's (2005, 82) characterisation of myth creation is generous in this way. For Salmon a myth is *any* mistaken theory that has been held as true and mythical objects are the postulates of these mistaken theories. But if the creationist is generous in this way she faces a problem; every distinct mistaken conception of an object can be taken as generative of a distinct mythical object. For example, when Hob and Nob characterise the target differently we may simply interpret them as having two distinct mythical objects in mind, posits of distinct mistaken theories. The creationist may still incorrectly predict that there is no intentional identity at (b).

But even if this particular problem for Salmon is avoided. The creationist does not appear to be out of the woods. She still needs to provide an assignment principle that will explain why particular empty intentional attitudes get assigned particular mythical objects. It looks as though considerations of fit alone cannot yield such a story and there appear to be no other explanatory resources available to explain intentional identity. Providing a creationist-friendly assignment principle is a

difficult task, especially given that the creationist ought to be generous about which mistaken conceptions give rise to an abstract object.

### 2.6.3. Refusing the challenge

There is another response to the problem that is still on the table. This option involves resisting the claim that we even need a theory of how attitudes get to be about the exotic objects they are about. One might claim, as some are tempted to when engaged in debates about the foundations of intentionality, that the aboutness relation that stands between attitudes and the objects that they are about is *primitive, brute* or in some other way unanalysable. In this way one might justify not facing up to the challenge of providing an assignment principle on the grounds that any attempt to provide an explanation or analysis of the relation will be bound to fail. I admit that this is a way of sidestepping the problem. However, I will note that this position is profoundly unsatisfying to those engaged in the debate concerning how intentional identity ought to be understood.

Another way of refusing the challenge involves claiming that there are facts concerning which attitudes are about the same exotic objects but that these facts are, in some crucial sense, unknowable. This claim amounts to view of intentional identity that is somewhat analogous to an epistemicist theory of vagueness. In particular, it resembles the kind of epistemicism defended by Sorensen (1988). If this kind of view is correct there is, once again, no reason to think that any theory concerning which objects get assigned to which attitudes could ever be adequate. This is another way of refusing the challenge which, though open, is unsatisfying; we ought not to give up looking for an explanatory theory of intentional identity quite so hastily.

### 2.6.4. Hybridising

If a transparent theory is right, intentional identity depends on the presence of an object that the relevant attitudes are about. So it is natural, when developing a transparent theory, to look for ways to link the *exotic objects themselves* to empty intentional states in a principled manner. But resources that link particular empty intentional states to exotic objects are not the only resources available to explain intentional identity. There are, after all, many non-transparent theories of intentional identity in the literature and they purport to explain intentional identity without insisting

on the presence of exotic objects.

Perry (2001, 14, 147-156) and Sainsbury (2005, ch. 7, 2010), for instance, claim that intentional identity is a matter of there being a causal chain that links the attitudes themselves together.<sup>10</sup> In scenario (a), for example, there is a causal chain that goes via the newspaper article that links Hob's belief with Nob's belief and this causal link is supposed to explain the fact that those beliefs have a common focus. Perry (2001, 150-160) and Sainsbury (2005, 75-77, 237-238, 2010) attempt to account for intentional identity without claiming that empty attitudes are, in any sense, *about* an object.

Others, like Dennett (1968) and King (1993, 76), attempt to account for intentional identity in terms of a shared conception of the target object. The idea is that if the agents characterise the target in a similar way then the empty attitudes get to have a common focus. Once again, these theories do not involve appeals to an object that the relevant attitudes are about.

Another sort of non-transparent theory is presented by Geach (1967, 627). For Geach, intentional identity is a matter of the empty attitudes in question involving the same *aspect* (a kind of Frege-style mode of presentation). Some aspects are aspects *of* something. When I have beliefs about London, for instance, I employ an aspect that is an aspect *of* London, the city itself. However, some aspects are not aspects *of* anything. If there are no witches, when I believe something about a particular witch, the witch aspect that is involved in my belief is not an aspect *of* a witch; there is nothing that satisfies the aspect. Geach's view is another example of a non-transparent account. It does not require the presence of exotic objects that the attitudes are about.

At this point it will be helpful to distinguish *transparent* resources from *non-transparent* resources. Non-transparent resources are the kinds of things that non-transparent theorists appeal to in their accounts of intentional identity: causal links between attitudes and agents, similarities between how the agents conceive of the object, which concepts agents employ, and so on. Non-transparent resources characteristically do not require that there be an object that the relevant attitudes are

---

<sup>10</sup> There are reasons to think that theories based entirely on causal links will not be adequate, see Edelberg (1992) and van Rooy (2000, 179 n.22).

about.<sup>11</sup>

Transparent resources, on the other hand, concern exotic objects directly. If particular exotic objects made a distinctive causal difference to concrete objects like Hob and Nob, we could appeal to these distinctive causal properties when assigning exotic objects to empty attitudes and this would be a transparent resource. Likewise, if some particular exotic objects were just more eligible to be the objects of attitudes than others, and this made a difference to which objects attitudes were about, then this would constitute a transparent resource that we could appeal to. Call transparent theories that involve only appeals to transparent resources, *pure transparent theories*.

One option still open to transparent theorists is to help themselves to some of these non-transparent resources. Transparent theorists might *hybridise* their theory. The idea is to pick the best non-transparent theory of intentional identity, whatever it might be, and appropriate the non-transparent resources that theory appeals to, in the service of a transparent account.<sup>12</sup> The hybridising strategy involves the transparent account of intentional identity *piggybacking* on a non-transparent account of intentional identity by appropriating its resources. Call transparent theories that involve an appeal to non-transparent resources, *hybrid theories*. This response may seem somewhat attractive but I will argue that making this move has profound implications for the prospects of a transparent account of intentional identity.

So this is how I see the dialectic as proceeding; the transparent theorists say that attitudes are g-related just in case there is an object that they are both about. I then claim that we want a story about *when* and *why* attitudes are about the same object. But when the transparent theorist attempts to tell this story just in terms of transparent resources, they run into the problem discussed in section 2.5. I argued that, in light of the poverty and abundance observations, transparent resources

---

<sup>11</sup> Of course, as we will see, they tend to be *compatible* with the presence of an object of this sort.

<sup>12</sup> Note what hybridisation will *not* yield. The addition of non-transparent resources will yield verdicts about which empty attitudes are about the same thing. It will not, in general, yield verdicts as to which exotic object this or that empty attitude is about. If we had an account of when and why exotic objects are assigned to particular attitudes, we would thereby have an account of when and why two or more attitudes are about the same object, but the opposite is not true.

will not yield an acceptable story about assignment in the hard cases. So the transparent theorists need to help themselves to some non-transparent resources.

But if the transparent theorist concedes in this way and moves from a pure transparent theory to a hybrid theory, it is very bad news for the transparent approach to intentional identity overall. The real work of explaining intentional identity in the hard cases is done by the non-transparent resources to which the hybrid theorist appeals. The hybrid theorist only gets to deliver the correct verdicts by appropriating the non-transparent theorist's resources and by mimicking the behaviour of a non-transparent theory. According to non-transparent theorists, non-transparent resources alone are enough to provide an explanation of intentional identity. The hybrid theorist can hardly disagree. The hybrid theorist cannot claim that no set of non-transparent resources could yield a good theory of intentional identity, on pain of not having a satisfactory set of non-transparent resources to appropriate. What is more, if the poverty observation and the abundance observation are correct, positing exotic objects yields no extra resources which the hybrid theory has over the corresponding non-transparent theory; adding on the objects on top of the non-transparent theory does not do any explanatory work. If this is right, it is not clear how intentional identity depends, in any deep way, on the presence of an object that the relevant attitudes are about. By hybridising their theory, the transparent theorist gives the game away.

## 2.7. Conclusion

Let us take stock. I have argued that if the abundance and poverty observations are correct, as the transparent theorist ought to think they are, then the transparent theorist cannot deliver the correct verdicts about intentional identity, at least if they limit themselves to transparent explanatory resources. For this reason we ought to reject transparent theories that limit themselves to transparent explanatory resources.

One seemingly attractive line of response is the hybridising move. This option involves the transparent theorist conceding that transparent resources are insufficient to yield an adequate account of intentional identity, but supplementing their theory by appealing to characteristically non-transparent resources. This move undermines the overall transparent strategy, as it involves admitting

that positing the objects does not give us any extra explanatory recourse when it comes to explaining intentional identity. Even if we are tempted by the hybridising view, we should not posit exotic objects in the service of explaining intentional identity.

One lesson that can be taken from this result is that everyone in the debate about intentional identity ought to care about which non-transparent theory of intentional identity is best. The transparent theorists ought to care because they need to decide how to hybridise their theory. Should they hybridise their theory by drawing on the work of Perry, Sainsbury, King or some other non-transparent theorist? The answer to this question will turn on which of the non-transparent theories comes out the strongest. Non-transparent theorists ought to care because they want their theories to be better than their non-transparent rivals.

We may have independent motivation to countenance exotic objects and include them in our theory of empty intentional attitudes. If we are going to be committed to exotic objects for some reason other than the need to explain intentional identity, we should expect them to show up when empty attitudes have a common focus. Another lesson that can be taken from the above discussion is that the only reason for adopting a hybrid transparent theory of intentional identity over the corresponding non-transparent alternative is that one is already committed to countenancing exotic objects for some other reason, since adding exotic objects does not help us explain intentional identity.

However, there are reasons why we might want to do without exotic objects altogether. Views that countenance exotic objects tend to inherit certain metaphysical puzzles concerning the nature of the entities in question. Russell (1905, 482-483) famously objected to non-existent objects on the grounds that they purportedly end up committing those who believe in them to inconsistencies and there is a large literature concerning how a defender of non-existent objects ought to avoid this pitfall. Quine (1948) objects to the idea of merely possible objects on the grounds that they are metaphysically dubious. There are also puzzles concerning how many of the relevant objects there are. For a discussion of some such puzzles see Nolan (1996), and Nolan and Sandgren (2014). Brock (2010) has recently launched a powerful attack on creationist approaches to the metaphysics of mythical objects. These are just some of the challenges and hazards that those who appeal to exotic objects face. If we can get away with leaving those exotic objects out of our theories we could avoid these complications altogether.

Those who do not appeal to exotic objects also have an advantage when it comes to explanatory and ontological simplicity. All else being equal, we ought to prefer a simpler explanation. All else being equal, we ought to prefer theories that are committed to fewer kinds of things. When comparing a hybrid transparent view with a non-transparent theory that employs the very same non-transparent resources the latter theory appears to have the advantage of explanatory and ontological simplicity.

Metaphysical posits like exotic objects earn their keep by helping to explain things. The above discussion suggests that positing exotic objects does not help us to explain intentional identity. It may be that countenancing such things is justified because it explains some other phenomenon. But if what I have said above is correct, there is at least one fewer thing that an appealing to exotic objects helps us explain. Exotic objects will have to earn their keep *somewhere else*. Arguments like those presented above chip away, bit by bit, at the motivations for countenancing exotic objects. If things continue like this, the costs of countenancing exotic objects may eventually outweigh the explanatory benefits they confer.



## CHAPTER 3: MORE THEORIES OF INTENTIONAL IDENTITY

### 3.1. Alternatives

In the previous chapter I considered the prospects of a transparent approach to intentional identity. In this chapter I will survey some of the principal non-transparent alternative theories. As I proceed through the survey I will raise complaints against these theories, arguing that they all have serious shortcomings of one sort or another.

For the transparent theorist, the g-relation is to be understood as an *about the same thing* relation. But this is only one way of understanding the g-relation. What other options are available? Some of these theories have been mentioned before in chapter two, but I will lay them out in a little more detail here.

### 3.2. Dennett's theory

Let us start with an account of intentional identity put forward by Dennett (1968, 336-338, 341). According to Dennett, there is intentional identity if, and only if, the agents share *all* beliefs regarding the putative target.

‘[F]ar from its being the case, as Geach contends, that no descriptions need be shared for us to speak of intentional identity, on the contrary, unless all descriptions are shared, the notion of the identity of intentionally inexistent objects dissolves into nonsense.’ (1968, 337)

This account is clearly inadequate. If it were correct there would be no intentional identity in cases in which the agents disagree about the putative target. If the agents disagree they will have different beliefs about the target and thus, by Dennett's lights, there would not be intentional identity. Consider (b) and (1), repeated here for ease of reference.

- (b) Hob and Nob live in the same village and read the same newspaper. The newspaper reports that a witch has been terrorising the village. Hob believes that the witch mentioned in the newspaper blighted Bob's mare. Nob thinks the witch mentioned in the newspaper did not blight Bob's mare but did kill Cob's sow. Witches do not exist.

- (1) Hob thinks a witch has blighted Bob's mare, and Nob believes she (the same witch) killed Cob's sow.

According to Dennett, (1) will be false at (b). This is the wrong result, (1) may be true at (b). This same problem will arise in all disagreement cases in which the attitudes are empty. For this reason Dennett's theory should be rejected.

The unattractiveness of this result can be made vivid by considering cases of disagreement in science. If we accept Dennett's view, we will be forced to claim that there are no cases of genuine disagreement in which the relevant attitudes are empty. For example, no two scientists have ever genuinely disagreed about phlogiston, black bile or Vulcan; they *always* had different entities in mind. Also, if it turns out that one or more of the posits of current physics do not exist, all the physicists who appear to be disagreeing about such entities today are not even talking about the same stuff. This is a very undesirable result.

### 3.3. The lazy theory

Next consider what is sometimes called *the lazy theory* or the *pronoun of laziness theory*. According to the lazy theory, the 'she' in sentences like (1) should be understood as what Geach (1967, 630) calls a 'pronoun of laziness'. Accordingly, the pronoun 'she' in (1) ought to be interpreted as standing for a definite description taken directly from (1)'s first conjunct. In line with the lazy theory, we might suggest either (3) or (4) as an analysis of (1).

- (3) Hob believes that a witch has blighted Bob's mare, and Nob believes that the witch that blighted Bob's mare killed Cob's sow.
- (4) Hob believes that a witch has blighted Bob's mare, and Nob believes that the witch that Hob believes blighted Bob's mare killed Cob's sow.

McKinsey (1986, 168) defends of one version of the lazy theory. McKinsey claims that (1) should be analysed as something like (3). Before considering the merits of (3) and (4) as analyses of (1) I should note that this sort of account is, at heart, an account of the semantics of intentional identity sentences rather than primarily a theory about intentional identity. It will, however, be instructive to note how (3) and (4) fail be good analyses of (1).

(3) is not a good analysis of (1). As Geach (1967, 630) correctly notes, (1) may be true even if Nob has no beliefs about Hob or about Bob's mare. We need only consider a variant on (a), repeated below for ease of reference, according to which it is stipulated that Nob has no beliefs concerning Hob or Bob's mare.

- (a) Hob and Nob live in the same village and read the same newspaper. The newspaper reports that a witch has been terrorising the village. Hob believes that a witch (the one mentioned in the newspaper) has blighted Bob's mare and Nob believes that the witch mentioned in the newspaper killed Cob's sow. Witches do not exist.

In such a scenario (1) may be true though (3) would be false.

There is another, more interesting, reason to reject (3) as an analysis of (1). The problem is easily understood if we consider scenario (b) again. (3) is false at (b) because the property mentioned in the first conjunct of (1) is the very feature of the witch about which Hob and Nob disagree. Since the move from (1) to (3) involved taking material from the first conjunct and using it to interpret the pronoun in the second conjunct, the account falters in cases in which the property mentioned in the first conjunct is disputed by the agents in question. The sort of lazy account that suggests (3) as an analysis of (1) is unduly beholden to the content of the first conjunct.

Now for the kind of lazy theory according to which we ought to analyse (1) as (4). Depending on how we read the second conjunct, this view would have the same problem as the other lazy theory concerning cases in which Nob does not believe anything whatsoever about Hob. There is a way of understanding the second conjunct of (4) according to which it requires that Nob has a belief about Hob and his beliefs. On this understanding this analysis runs into one of the same problems as the other lazy view.

The second conjunct of (4) could alternatively be understood as saying that Nob has a belief that is, in fact, about the witch that Hob's belief is about, whether or not Nob has any beliefs about Hob. But if we understand (4) in this way, it turns out not to be an analysis of intentional identity; (4) *contains* an ascription of intentional identity. To evaluate (4) we need to know whether the belief ascribed to Nob in the second conjunct really is about the same witch as Hob's belief is. Even if (4) is true in all the same scenarios as (1) is, the claim that (4) is the deep structure of (1) would not get us anywhere if the project is to find an account of intentional identity. Without an

independent understanding of intentional identity we ought to be just as confused about this reading of (4) as we are about (1), at least as regards intentional identity.

### 3.4. King's theory

King presents a theory that involves analysing (1) in roughly the same way as the lazy theorist who presents (3) as an analysis of (1). King (1993, 65) takes issue with (3) as analyses of (1) because he resists the interpretation of the pronoun. King presents an alternative account according to which the pronoun in (1) is, instead, interpreted as what King calls a 'context dependent quantifier'. The details of King's view will not concern us here; it is enough to note that according to King's account, (1) is equivalent to (5).

- (5) Hob believes that there exists a witch who has blighted Bob's mare, and Nob believes that there exists a witch who blighted Bob's mare and killed Cob's sow.

As King (1993, 76) admits, his account does mean that (1) is equivalent to (5) and that the truth of (5) entails that Nob has beliefs about Bob's mare. So if we adopt (5) as analysis of (1) we are committed to (1) having that entailment. But we saw above that (1) may be true while (5) is false, in cases in which Nob has no beliefs about Hob or Bob's mare.

King attempts to show that this result is not as undesirable as one might have thought. In response to the charge that this entailment flies in the face of our judgments about the truth conditions of (1), he claims that these judgments should be resisted. King (1993, 77) cites with approval similar attempts to explain away the intuitive costs of theories of belief according to which to believe that Mark Twain is a great writer *just is* to believe that Samuel Clemens is a great writer. King's defence of his view is not convincing and since it delivers wrong verdicts about simple cases of intentional identity, we ought to reject it.

### 3.5. Manning's theory

Another account of intentional identity has been presented recently by Manning (2015). According to Manning intentional identity sentences like (1) are always false when there are no witches. Manning's idea is, roughly, that (1) commits the utterer of (1) to the existence of the witch. For this reason, if Manning is right, sentences like (1) are bound to be false when there are no witches.

Intentional identity sentences ascribe identity and there is, if Manning is right, no identity without an entity. Since there is no entity in scenarios like (a), (1) cannot be true at (a). Braun (2012) also suggests that it is a mistake to think that there is any genuine intentional identity reading of (1) that gets to be true in cases like (a).

This move is desperate. Intuitively, (1) can be true in cases in which the putative target of the respective beliefs is missing. What is more, given the ubiquity of intentional identity, this move commits those who make it to the claim that, if there are no moral properties, no one, not even those who have worked on ethics most of their lives, has ever had two or more intentional attitudes about moral properties. Another commitment is that, if there is no god, no two believers ever had beliefs about the same god. Perhaps more worryingly, if some of the posits of contemporary physics turns out to not exist, all the present day physics do not have attitudes directed at the same stuff. For these reasons, this account of intentional identity is very unsatisfactory, it is a surrender to the puzzling features of the phenomenon.

Manning does attempt to explain why we might have mistakenly thought that sentences like (1) get to be true when there are no witches, by appealing to the *pragmatics* of (1) (what is conveyed by (1) but not literally expressed). The idea is that though (1) is false, it is sometimes appropriate to utter because it pragmatically conveys some information that is not literally expressed by the sentence. I don't find this kind of response convincing. Sentences like (1) do seem to be true at (a) and it seems, at the very least, a major cost of a view if it says that they are not. What is more, the assignment of truth values to intentional identity sentences seems to be systematic in a way that is incongruent with the kind of explanation Manning provides.

### 3.6. Geach's theory and Sainsbury's theory

Next up are theories of intentional identity that I will call *common element theories*. The idea behind common element theories is that attitudes being g-related depends on there being some representational entity, a concept, mode of presentation or some such thing, which the attitudes in question both involve. The fact that the attitudes employ the same representational entity is supposed to explain the presence of intentional identity.

For example, Geach (1976, 314-317) appeals to entities called *aspects*. Aspects are supposed to be

a kind of mode of presentation *of* things. For example, Venus may appear in the morning or in the evening. It has a morning aspect and an evening aspect but both of these aspects are aspects *of* the same thing, namely Venus. Geach's aspects are supposed to be close cousins of Frege's (1892) senses.

According to Geach, aspects feature in the content of empty attitudes. An aspect need not be an aspect *of* anything, for it to feature in the content of an intentional attitude. A witch aspect need not be an aspect of a witch. Consider a child who has the false belief that there is a monster under her bed and that it is hairy. In this case the monster aspect is not an aspect of anything, there is no monster, and yet the aspect is crucial for capturing what the child believes. With this aspect machinery in place, Geach is in a position to present his account of intentional identity; empty attitudes are g-related if, and only if, the relevant empty attitudes involve the very same aspect. What is more, the fact that the attitudes in question involve the same aspect is supposed to *explain* the fact that they are g-related.

Sainsbury (2005, ch. 7, 2010) defends another kind of common element theory. Sainsbury's view involves what he calls *individual concepts*. Sainsbury's individual concepts may be empty, there may be nothing that a concept picks out, and may yet feature in the correct model of empty intentional attitudes.<sup>13</sup> The child can employ the concept of the hairy monster, and it can feature in the correct model of her attitudes, even if there is no such monster. According to Sainsbury, empty attitudes are g-related if, and only if, they involve the same concept. Again, the fact that they involve the same concept is supposed to explain the presence of the g-relation and, in turn, the truth of certain intentional identity sentences.

### 3.7. Glick's theory and van Rooy's theory

The class of theories I will discuss next are what I will call *counterpart theories*. Like the defenders of transparent theories, the counterpart theorists claim that empty attitudes are about objects of some stripe. But unlike the transparent theorists, the counterpart theorists do not insist that for

---

<sup>13</sup> Sainsbury does not think that the differences in which concepts are involved in intentional attitudes necessarily makes a difference to the *content* of the attitudes. Sainsbury wants to reserve the term 'content' for the things that the concepts involved in an attitude pick out.

attitudes to be g-related they need to be about the very same object. For the counterpart theorist the objects that the respective attitudes are about need only be counterparts. If counterpart theory is right, there is intentional identity just in case the intentional objects that the respective attitudes are about are counterparts. What is more, these objects being counterparts is supposed to explain the presence of intentional identity. Van Rooy (2000, 170-178) endorses a theory of this kind. More recently, Glick (2012) has developed a counterpart theoretic account of intentional identity couched in the familiar possible worlds account of mental content. According to Glick, apparently empty beliefs are about merely possible objects. I will have more to say concerning Glick's view below.

### 3.8. Complaint 1: Incompleteness

At this point I would like to make a complaint. Some of the views discussed above are incomplete; they leave central questions about intentional identity unanswered. Call this *the incompleteness complaint*.

To take an example, Geach's theory is, as he admits (1976, 318), incomplete in that he is not explicit about *when* and *why* two attitudes involve the same aspect. He does not tell us enough about how to individuate aspects and how they get to be involved in an agent's attitudes for us to be able to judge if two empty attitudes do involve the same aspect, even in the simplest intentional identity cases. We are told that the g-relation depends on identity of aspect but we are not told how to identify aspects. Geach does put some constraints on how aspects ought to be individuated, in particular they are *not* to be individuated by the descriptions that the agents associate with the putative target of their attitude. But this is consistent with a large range of ways of filling in the details of the theory. Without this part of the picture it is not clear what verdicts about intentional identity Geach's theory delivers in particular cases. We need what we might call an *individuation principle* for aspects before we are able to evaluate Geach's theory in a satisfactory way.

The same complaint applies to the counterpart-theoretic approaches presented by Glick and van Rooy. We are told that intentional identity is a matter of certain objects being counterparts. But what does it take for possible objects or belief objects to be counterparts? Glick (2012, 390, 393) admits that he does not provide detailed conditions on how we should understand the relevant

counterpart relations. The relation in question is supposed to be a relation of comparative similarity, but we are only given minimal constraints on which respects of similarity are relevant. He also claims that causal factors influence our judgements about intentional identity ‘even if it is not clear exactly how’ (Glick 2012, 393). Van Rooy (2000, 172) says that standing in the relation in question is, in most cases, a matter of the empty beliefs having a common causal source. He admits, however, that sometimes there is intentional identity in the absence of any such causal link (van Rooy 2000, 179 n.22). It is again left open *how* and *when* the intentional objects in question are related in the appropriate way. These theories are missing what we might call a *linking principle* that will dictate when two objects are counterparts in the appropriate sense. While it is true that both Glick (2012, 391-393) and van Rooy (2000, 172-173) claim that intentional identity has something to do with causal links between the relevant empty beliefs, at least sometimes, neither are explicit or precise about the role of the causal links in their respective theories of intentional identity.

It is also worth noting that I levelled this same complaint, in a slightly different guise, in chapter 2 against transparent theories such as those presented by Salmon (2005, 105–108), Parsons (1974, 577–578) and Saarinen (1978, 207–210). As alluded to in chapter 2, these theorists do not provide a theory of how attitudes come to be about what they are about or a theory of when and why intentional attitudes are about the very same object. Since, according to Salmon and Parsons, g-relatedness is a matter of beliefs being about the same thing, if we aren’t told when two beliefs are about the same thing we cannot judge when beliefs are g-related. In this way it is unclear *when* two attitudes are, by their lights, g-related. What we need is a story about *how* empty beliefs come to be about the objects they are about or at least a theory of when they are about the same thing. I argued in chapter 2 that there does not appear to be any good way of giving a theory of this kind, if the transparent theorist confines themselves to transparent recourses, but even if this argument lapses these transparent theories are still vulnerable to the incompleteness complaint.

Glick (2012, 390), pre-empting the charge of incompleteness, claims that incompleteness is not a reason to object to *his* account of intentional identity since all of its rivals in the literature are also sketchy and incomplete. This remark is rather strange, there were more complete theories of intentional identity to be found in the literature at the time when Glick’s paper was published. I will discuss some such theories in the next section. I will argue that those theories should be rejected



for other reasons, but at least these theories provide a more complete and precise account of intentional identity than the theories, like Glick's, that are more open to the incompleteness complaint.

These theories are incomplete and, all else being equal, we ought to prefer more complete theories over less complete ones. This is a serious complaint against some of these theories *as they stand*. If we want an adequate account of intentional identity, the defenders of these views must fill in the details of their theories more than they have done. This complaint can be thought of as an invitation for defenders of incomplete theories to say more. Still, the complaint ought to be given its proper weight.

### 3.9. The causal link approach

Those who *have* provided a more complete theory tend to adopt what I call the *causal link approach*. This approach is based on what I call the *causal link thesis* which says that intentional identity is primarily a matter of the beliefs being linked together by an actual causal chain and that this chain makes for intentional identity in a way that is not, in general, dependent on how subjects take the putative target of their beliefs to be. Sainsbury (2005, ch. 7, 2010, 2011, 105-106), Perry (2001, 14, 147-156), Friend (2014), Glick (2012, 391-393), Zimmerman (1999) and van Rooy (2000, 172-173) all adopt the causal link approach, though they implement the approach in different ways, arriving at slightly different theories of intentional identity. Lewis (1986, 34) also hints at something like the causal link thesis, though he does not develop or defend it in any detail.

A natural thought when confronted with (a) is that the presence of the newspaper article is crucial for explaining the fact that Nob's belief is g-related to Hob's and thus crucial for the truth of (1). Both Hob and Nob's beliefs are linked together by a casual chain that involves the newspaper article as a causal source of both empty beliefs. Intuitively, it seems as if the presence of this causal chain linking these two empty beliefs together is a central part of the explanation of the presence of intentional identity in scenario (a). This suggests a general approach to intentional identity, *the causal link approach*, based around the causal link thesis.

For example, according to Perry (2001, 14, 147-156), there is intentional identity when there is an

*inter-subjective causal network* that links the relevant beliefs together.<sup>14</sup> Attitudes are g-related when they are ‘supported’ by the same such network. Inter-subjective networks are comprised of agents’ uses of a term, connected together by a chain of communicative and deferential interactions between agents. Each network has exactly one causal origin and networks are individuated by their causal origin. If Perry is right there is intentional identity just in case the intentional states are linked together by a causal network of this sort. The presence of a network is supposed to explain the presence of intentional identity.

For Sainsbury, the causal link between the attitudes gives rise to intentional identity in a different way. Sainsbury, it will be remembered, ties g-relatedness to identity of concept employed. The causal links between the two empty intentional states come in when we ask *how* concepts are individuated and when two or more attitudes involve the same concept. He develops a theory of how concepts are to be individuated in joint work with Tye (2011, 2012). Note that, though the theory of concepts is developed in joint work with Tye, it is only Sainsbury (2010) that employs this theory of concepts in the interests of accounting for intentional identity. The idea is that there are causal chains of deferential uses of a concept leading back to its causal origin or ‘originating use’. For every concept there is just one originating use and concepts are wholly individuated by their originating use; concepts are identical if, and only if, they have the same originating use (Sainsbury and Tye 2011, 105-106). For Sainsbury, the causal links make a difference to intentional identity *indirectly* by making a difference to concept individuation.

van Rooy and Zimmermann (1996), Zimmermann (1999), Perry (2001, 14, 147-156), Glick (2012, 391-393) make a similar move though in a less full blooded way. As noted above, Glick and van Rooy both claim that causal links between the relevant empty beliefs have something to do with

---

<sup>14</sup> It is not clear how Perry intends to handle intra-personal intentional identity cases. Presumably there are cases of intentional identity in which there is no intersubjective network linking the beliefs together. I may form beliefs about the monster under my bed and have a belief about that monster on both Monday and Tuesday even if there is no monster. What is more, this may occur even if I never tell anyone about my monster beliefs. I will leave this question open, but the defender of Perry’s account will need to find a way of explaining this sort of intentional identity that does *not* involve inter-subjective networks.

intentional identity though they are not explicit about this relationship. Van Rooy (2000, 173) suggests, taking his cue from Zimmerman (1999), that it is a necessary condition for intentional identity, though he seems to go back on this claim in a footnote (van Rooy 2000, 179 n.22).

Theories based on the causal link thesis are better than some of their rivals. They can deliver the correct verdict that (1) may be true at (a), account for intentional distinctness (cases in which empty attitudes are *not* g-related) and allow intentional identity in spite of disagreement concerning the target.

Consider again scenario (a). In (a) Hob's belief and Nob's belief are causally downstream of the same newspaper. If the causal link thesis is right, this fact could explain the presence of intentional identity. So if the causal thesis is true and reflected in our theory of intentional identity, we can deliver the correct verdict about (a).

Next consider (b). At (b), just as at (a), Hob's belief and Nob's belief are causally downstream of the same newspaper article so if the causal link thesis is true and reflected in our theory of intentional identity then we may deliver the correct verdict that there is intentional identity at (b).

This fact reveals a crucial feature of causal link theories, they all involve linking intentional identity to facts that need not be reflected in how the agents take the putative target of their beliefs to be. The fact that Hob and Nob disagree about the target at (b) need not, according to defenders of the causal link approach, prevent the attitudes in question from being g-related. The causal link is present whether or not that fact is reflected in either agent's characterisation of the target.

It is worth noting, at this point, that the causal link thesis comes in varying strengths. We can distinguish what we might call *the strong causal link thesis* from weaker versions of the causal link thesis. The strong causal link thesis says that the causal link in question gives rise to intentional identity *on its own*, independently of how the relevant agent takes the putative target of the attitude to be. The strong causal link thesis implies that how the believer characterises the target of their beliefs does not make any difference to whether or not there is intentional identity. If this thesis is correct then no amount of disagreement between the two believers could ever be a threat to intentional identity, given the presence of the causal link.

Weaker versions of the causal link thesis say that the believer's characterisation of the target entity plays some role in giving rise to intentional identity. The characterisation and the causal link, in

combination, give rise to intentional identity; the causal link and the agent's beliefs about the target of their beliefs can be traded off. If this is true then *some* differences in characterisation threaten intentional identity. At any rate, in cases like (b) the disagreement may be sufficiently small so as not to threaten intentional identity; the characterisations are similar enough such that, once the causal link has done its work, there is still intentional identity. Weak causal link theses leave open *how* the characterisations and the causal link combine to give rise to intentional identity and different theorists may fill in these details as they see fit.

Finally, recall (c).

- (c) Hob, a butcher in Arkansas, believes that a witch lurks under his bed and she has blighted the mare of his neighbour Bob. Nob, a baker in Ukraine, believes that a witch lurks under his (Nob's) bed and that she killed the sow of Nob's neighbour Cob. Hob and Nob have never met and there has been no relevant interaction between their social circles. Witches do not exist.

At (c) there is, plausibly, no causal chain of the appropriate kind linking the two beliefs together, no newspaper, no chain of communication etc. The relevant causal link is absent. In this way causal link theories can deliver the correct verdict that Nob's belief and Hob's belief are not g-related in scenario (c).

### 3.10. Complaint 2: Which causal links?

I am now in a position to make another complaint, aimed at those who adopt the causal link approach.

The complaint is that those who adopt the causal link approach do not provide an acceptable account of *which* causal chains matter for intentional identity. In scenario (c), both Hob's belief and Nob's belief are causally downstream of the big bang. But this does not mean that (1) gets to be true at (c); only *some* causal chains should be taken to make for the relevant g-relation. Those who appeal to the causal link thesis need a way of distinguishing the causal chains that give rise to intentional identity from those that do not. Kroon (1987, 6-9, 2009, 152) presses a challenge of this kind against causal theories of reference. Call this *the causal imprecision complaint*.

One obvious way of attempting to make this distinction is by distinguishing the causal links that

involve communicative, deferential interactions between agents from those that do not, claiming that if a causal chain gives rise to intentional identity it must involve communicative interactions between agents. The causal chain that links Hob and Nob together via the newspaper article in scenarios like (a), involves communicative and causally efficacious interactions between agents. On the other hand, the causal chain that links Hob and Nob together via some cosmic events in the distant past, has parts that do not involve communication or deference. Sainsbury (2011, 102-105) and Perry (2001, 14, 147-156) are both committed to the claim that the only causal chains that make for g-relatedness are ones involving communicative interactions between agents.

To see why this will not work, consider scenario (e), a variant on a case presented by Edelberg (1992, 573-575).

- (e) Floyd stages a fake car accident. He smashes the windshield and makes it look like the driver sailed through it on impact. He tows the car to another location where he splashes ketchup on the ground in front of the car. After a few minutes Tanya walks by and is taken in. She comes to believe that the victim of the crash went through the windshield. Ten minutes later, after Tanya has gone, Hank comes by and is fooled by the hoax. He also thinks the victim of the crash went through the windshield.

Tanya and Hank's beliefs directed at the victim have a common focus, even though there is no such person. (6) is intuitively true in scenario (e).

- (6) Tanya believes the victim of the car accident went through the windshield, Hank believes she (the same victim) went through the windshield.

Yet there were no communicative interactions between Tanya and Hank that involved the right kind of deference, and no third party to form an intermediate link in the communicative, deferential chain. So the relevant distinction cannot be made in this way on pain of yielding the wrong verdict about (e) and cases like it.

The reader may object that in scenario (e), Floyd himself somehow forms the intermediate link in the communicative chain and that, as a result, a communicative chain-based approach might be able to explain intentional identity in this case. This objection is easily avoided. There are cases like (e) in which there is no victim and the misunderstanding was not intentional. For instance, we can alter the case so the car accidentally fell off the back of a truck and landed in such a way so as

to make it appear as if there was a victim who went through the windshield. In such a case, there may be intentional identity in the absence of the kind of communicative chain that may exist in scenario (e).

But even if we were able to make the distinction in this way, we would still not be out of the woods. Even if we only consider causal chains involving communicative, deferential interactions between agents we still need a way of identifying the communicative interactions that give rise to intentional identity. Communicative interactions between agents are complex and multi-faceted. There are all sorts of communicative and deferential interactions that agents are involved in, and all sorts of corresponding causal chains. So the defender of the causal link approach is once again faced with a tricky question: which causal chains make for intentional identity? This constitutes a challenge that those who adopt the causal link approach have yet to adequately face.

### 3.11. Complaint 3: Causal links are not necessary for intentional identity

There is another complaint I want to make against causal link theories. The complaint applies to theories, like those of Sainsbury and Perry, which say that the presence of a causal link of the appropriate kind is *necessary* for intentional identity. Call this *the necessity complaint*.

There are cases of intentional identity in which the relevant kind of causal link is absent. Consider a case in which two mathematicians, call them Leibniz and Newton, have beliefs that target the same mathematical object. Suppose Leibniz and Newton, quite independently of one another, both come up with the idea of a mathematical function *integration* (though they use different words to refer to it) that plays almost exactly the same role in their respective mathematical reasoning. Suppose further that they are mathematical Platonists; they think that mathematical objects are abstracta that exist in Plato's heaven and that they are mistaken, there are no mathematical objects. It appears as if, in such a case, there is no causal link of the appropriate kind between the two beliefs since they came up with the idea independently. Yet it seems as if their beliefs may be g-related; their beliefs have a common focus in being directed at a particular function. Both Edelberg (1992) and van Rooy (2000, 179 n.22) discuss cases like these and conclude that there must be more to intentional identity than the presence of causal links between beliefs.

There are presumably many more cases of this sort. Most will accept that beliefs *sometimes* get to

be about particular objects by description, by means of the believer listing some of the putative target's properties, and not by means of a causal chain. If this is right then there will be cases in which beliefs have a common focus, in the relevant sense, where there is no causal link of the right sort binding them together.

One might object that there is some sort of common history in the Leibniz/Newton case, traditions of defining mathematical functions for instance. This may be true, but the point still stands. Consider a similar case in which Leibniz and Newton are on opposite sides of the galaxy and their communities have not interacted at all. In such a case, their attitudes concerning the integration function may be g-related even if there is no relevant common causal history of mathematical practice.

The presence of a causal link of the relevant sort between beliefs is *not*, in general, necessary for intentional identity. This result has two important implications. The first is that those, like Sainsbury (2005, ch. 7, 2010, 2011, 105-106) and Perry (2001, 14, 147-156), who claim that the presence of a causal link of the appropriate kind is, in general, necessary for intentional identity are mistaken. The second, more general, implication is that when it comes to providing an account of intentional identity, causal links cannot be the whole story.

### 3.12. An extra challenge

Some theories in philosophy of mind and philosophy of language involve appeals to *linguistic communities* of some stripe. A well-known example is the view, famously defended by Burge (1986, 1979), that what attitudes an agent has sometimes depends on the thought and talk of other agents in their linguistic community. More locally, some theories of intentional identity also appeal to thought and talk in the relevant linguistic communities. For Perry (2001, 14, 147-156), intentional identity depends on the properties of intersubjective networks of agents. These networks can be understood as communities of some sort. For Sainsbury and Tye (2011, 2012, 21-22, 24, 43), agents are able to latch onto shared concepts used by their linguistic community. In each case the linguistic community, what they say and how they behave, gets to make a difference to attitudes and, in turn, to intentional identity. Just as those who appeal to causal links to explain intentional identity need to say which causal links do the work, those who appeal to linguistic communities

need to say which communities make the difference. They need to say how communities are individuated and what difference they make. Just as there are a whole lot of causal links, there are all sorts of ways of delineating the relevant community that will make a difference to the verdicts the resulting theory delivers regarding intentional identity.

Which community is relevant? How should we delineate the linguistic communities in question? How do we decide which agents are members of the relevant community and which ones are not? These questions are not often answered precisely. Philosophers who employ the notion of the linguistic community in their theories, in whatever capacity, must present some principled account of which communities matter and why before their theories can be considered fully adequate.

### 3.13. Complaint 4: Non-uniformity

As discussed in chapter one, all else being equal we ought to prefer a theory of intentional identity that handles intentional identity in cases involving empty intentional attitudes and intentional identity in cases involving non-empty attitudes in a uniform way. Some of the theories above deliver a non-uniform treatment of intentional identity. This leaves them open to what I call *the non-uniformity complaint*.

Geach and Sainsbury treat the empty and non-empty cases differently. For Sainsbury and Geach, when beliefs are non-empty, having a common focus is a matter of the presence of an object that both beliefs are about; it is a matter of common reference. For instance, beliefs directed at Hesperus and beliefs directed at Phosphorous get to be about the same thing in virtue of their being about the same object, Venus. On the other hand, when the attitudes are empty they are ‘about the same thing’ in virtue of involving a common aspect or individual concept. In this way, Sainsbury and Geach handle empty and non-empty cases of intentional identity in fundamentally different ways. In fact, Sainsbury uses two different terms. He uses ‘internal singularity’ to refer to the kind of common focus empty intentional attitudes have and ‘external singularity’ to refer to the kind of common focus that non-empty intentional attitudes have. Any theory that, like Sainsbury’s and Geach’s requires a non-uniform treatment of intentional identity in this way has a strike against it.

### 3.14. Complaint 5: Symmetry



Some cases presented by Edelberg (1986, 1992) will allow me to make another complaint. Edelberg's cases provide grounds for another complaint against some of the theories discussed above; they are open to what I will call the *symmetry complaint*.

Consider scenario (f), inspired by Edelberg's (1995, 317) case involving two detectives.

- (f) Hob and Nob are partners in a witch-hunting business. They investigate the suspicious illness of Bob's mare. They both conclude that Bob's mare was blighted by a witch who was working alone. The next day they investigate the suspicious death of Cob's sow. At this stage a disagreement arises between Hob and Nob. Nob thinks one witch both blighted Bob's mare *and* killed Cob's sow. Hob thinks one witch blighted Bob's mare and a different witch killed Cob's sow. Witches do not exist.

On at least one reading of (1) and (7), (1) is true in scenario (f) though (7) is false.

- (1) Hob thinks a witch has blighted Bob's mare, and Nob believes she (the same witch) killed Cob's sow.  
(7) Nob thinks a witch has killed Cob's sow, and Hob believes she (the same witch) blighted Bob's mare.

The theories of Salmon, Parsons, Geach and Sainsbury deliver incorrect verdicts about (1) and (7).

Take, for instance, Salmon's theory. Salmon (2005, 105) presents (8) as an analysis of (1).

- (8) For some mythical witch *x*, Hob believes *x* blighted Bob's mare, and Nob believes *x* killed Cob's sow.

This lines up with Salmon's view that there is intentional identity just in case there is an object such that both beliefs are about that object. In line with this strategy, Salmon would, presumably, suggest (9) as analysis of (7).

- (9) For some mythical witch *x*, Nob believes *x* killed Cob's sow, and Hob believes *x* blighted Bob's mare.

The truth of (8) entails the truth of (9). But the truth of (1) does not entail the truth of (7); (1) but not (7) is true in scenario (f). So Salmon's theory of intentional identity must be inadequate. There must be some difference between (1) and (7) that is washed out when we adopt a Salmon-style

analysis of intentional identity. A similar problem will arise for Parsons (1974, 577–578), who presents a structurally similar proposal involving non-existent objects instead of abstract objects.

This should not really be surprising. The g-relation is, for Salmon and Parsons, an ‘about the same thing’ relation and if x is identical to y then y is identical to x. So, by Salmon’s and Parsons’ lights, if a pair of beliefs is about the same thing, they will stand in the g-relation, if they are not then they will not. In this way, defenders of a transparent approach will not leave room for (1) to be true while (7) is false.

The same sort of problem will arise for Geach and Sainsbury. Recall that Geach and Sainsbury tie the g-relation to an identity of concept or aspect. The g-relation is supposed to be an ‘involves the same concept’ relation or an ‘involves the same aspect’ relation. Either Hob’s belief and Nob’s belief involve the same aspect or concept, in which case (1) and (7) will both be true, or they do not, in which case (1) and (7) will both be false. This is the wrong result; (1) is true but (7) is false in scenario (f). Geach and Sainsbury do not deliver the result that (1) and (7) can differ in truth value.

The theories of Salmon, Parsons, Geach and Sainsbury do not have the resources to handle Edelberg-style cases; they do not leave room for (1) but not (7) to be true in scenarios like (f). This is a major shortcoming of these theories. What is more, it is not clear how one might attempt to adjust the theories so as to allow them to handle Edelberg-style cases while retaining the spirit of the original proposals.

### 3.15. Glick’s strategy

Glick (2012, 395-396) proposes a method for handling Edelberg-style cases, couched within his counterpart theoretic approach to intentional identity. According to Glick, attitudes are g-related just in case the merely possible objects that they are about are counterparts. For now we will concern ourselves only with beliefs and no other kinds of attitudes. Glick adopts the familiar model of an agent’s beliefs in which they are represented by a set of possible worlds. A world is a member of the set of an agent’s belief worlds if, and only if, nothing the agent believes rules out that the world is actual. In these possible worlds there are possible objects that we can say that beliefs are

‘about’ and these object stand in counterpart relations. Glick does not appeal to any identity relations that stand between things at different worlds. The crucial relations are counterpart relations; relations of comparative similarity between possible objects in contextually salient respects. As Glick admits, there are many different counterpart relations that correspond to different relations of comparative similarity, governed by different standards of similarity. Glick uses this feature to explain Edelberg’s cases. The idea is that *which* counterpart relation is invoked by a given intentional identity sentence depends on the linguistic context of the anaphoric pronoun (in (1) the pronoun is ‘she’) and which properties are especially contextually salient.

According to Glick, there are a whole series of g-relations corresponding to the different counterpart relations that, in turn, correspond to various relations of comparative similarity. Since two objects may be counterparts in one sense while not being counterparts in another sense and which counterpart relation is relevant for evaluating a given intentional identity sentence depends on the linguistic context of the pronoun, different intentional identity sentences can differ in truth conditions, even if they concern the very same attitudes.

Let me illustrate how Glick’s proposal works by considering how it delivers the right verdicts about (1) and (7) at scenario (f). All of Hob’s belief worlds contain two witches, one who blighted Bob’s mare and the other who killed Cob’s sow.<sup>15</sup> The first conjunct of (1), ‘Hob thinks a witch has blighted Bob’s mare’, comes out true at (f).

To evaluate the second conjunct ‘Nob believes she (the same witch) killed Cob’s sow’ we need to find objects in Nob’s belief worlds that are counterparts of the objects that blighted Bob’s mare in Hob’s belief worlds and check to see whether they killed Cob’s sow. According to Glick, the counterpart relation invoked is a relation of comparative similarity especially *in particularly contextually salient respects*. Glick says the property of having blighted Bob’s mare is salient in this context, since it was featured in the first conjunct. So we must look for objects in Nob’s belief worlds that resemble the objects that blighted Bob’s mare in Hob’s belief worlds *especially with respect to having blighted Bob’s mare* and check if they killed Cob’s sow. The objects in Nob’s belief worlds that are most similar in this way *did* kill Cob’s sow, so the second conjunct of (1)

---

<sup>15</sup> Strictly speaking, of course, they blighted and killed *counterparts* of Bob’s mare and Cob’s sow; we may harmlessly ignore this complication.

also gets to be true at (f). So far so satisfactory.

Next consider (7). All of Nob's belief worlds contain a witch who blighted Bob's mare and also killed Cob's sow. The first conjunct of (7) 'Nob thinks a witch killed Cob's sow' comes out true at (f).

To evaluate the second conjunct 'Hob believes she (the same witch) blighted Bob's mare' we need to find objects in Hob's belief worlds that are counterparts of the objects that killed Cob's sow in Nob's belief worlds and check to see whether they blighted Bob's mare. According to Glick, the property of having killed Cob's sow is salient in this context, since it was featured in the first conjunct. So we must look for objects in Hob's belief worlds that resemble the objects that killed Cob's sow in Nob's belief worlds *especially with respect to having killed Cob's sow* and check if they blighted Bob's mare. The objects in Hob's belief worlds that are most similar in this way *did not* blight Bob's mare, since in all of Hob's belief worlds the witch that blighted Bob's mare is distinct from the one that killed Cob's sow, so the second conjunct of (7) gets to be false at (f) and, in turn, so does (7) taken as a whole.

Glick's strategy neatly predicts the correct truth values of (1) and (7) in scenario (f). It is, however, limited in one crucial respect. Recall scenario (b) in which Hob and Nob disagree about whether or not the witch blighted Bob's mare. Hob thinks she did, Nob thinks she didn't. This disagreement is not, it seems, a threat to the presence of intentional identity, (1) is true at (b). But if we apply the Glick strategy we get the opposite result.

At scenario (b) all of Hob's belief worlds contain a witch that blighted Bob's mare and all of Nob's belief worlds contain a witch that killed Cob's sow but *did not* blight Bob's mare. The first conjunct 'Hob thinks a witch has blighted Bob's mare' gets to be true at (b).

To evaluate the second conjunct 'Nob believes she killed Cob's sow' we need to find objects in Nob's belief worlds that are counterparts of the objects that blighted Bob's mare in Hob's belief worlds and check to see whether they killed Cob's sow. Glick says the property of having blighted Bob's mare is salient in this context, since it was featured in the first conjunct. So we must look for objects in Nob's belief worlds that resemble the objects that blighted Bob's mare in Hob's belief worlds *especially with respect to having blighted Bob's mare* and check if they killed Cob's sow. However, in Nob's belief worlds the witch that killed Cob's sow *did not* blight Bob's mare.

So Glick's strategy leads us to deliver the incorrect verdict that (1) is *false* at (b).

Another way to bring out this difficulty is to consider a sentence like (10).

(10) Hob thinks a witch blighted Bob's mare, and Nob believes she (the same witch) did not blight Bob's mare.

(10) will sometimes be true. But if we employ Glick's strategy we will never be able to make this true, unless we adopt a different understanding of contextual salience. The agents in question disagree about the features ascribed to the witch in the belief mentioned in the first conjunct. Glick's strategy, as it is stated, can only be usefully applied to a certain proper subset of intentional identity cases.

It is worth noting that the lazy theory according to which (3) is supposed to be an analysis of (1) will also run into this same problem. Following the spirit of the lazy theory we might present (11) as an analysis of (10).

(11) Hob thinks a witch blighted Bob's mare, and Nob believes that the witch that blighted Bob's mare did not blight Bob's mare.

But (11) can be false while (10) is true, for instance in scenario (b). This sort of lazy theorist and Glick both make themselves unduly beholden to the content of the belief ascribed in the first conjunct. This means that cases like (b) and sentences like (10) make trouble for those views.

What kind of counterpart relation would need to be invoked if Glick's theory is to deliver the correct verdicts about cases like (b)? We might adjust the invoked counterpart relation so that it corresponds to a relation of comparative similarity with respect to being the thing that Hob *thinks* blighted Bob's mare. But if this is the intended counterpart relation then there is trouble. We have to know what it takes to be the thing that Hob thinks blighted Bob's mare before we can evaluate the intentional identity claim. That is, we need an independent grip on intentional identity in order to apply Glick's strategy. This is no help. We wanted to know about intentional identity, when and why beliefs are g-related. All we are told is that they are g-related when they have a common focus, which gets us nowhere. Just like the lazy theory that involves presenting (4) as an analysis of (1) is circular, Glick's theory is also circular, if the counterpart relation invoked is adjusted in this way.

Glick's strategy is either incomplete, in that it cannot be properly applied to some cases of intentional identity or it is circular. Either way, Glick's strategy leaves something to be desired, and we had better keep our eyes peeled for ways of avoiding these pitfalls and for a better way of handling Edelberg-style cases.

It is also worth noting that even if this part of Glick's picture is found wanting, his general approach to intentional identity need not be threatened. This part of his picture primarily concerns how intentional identity is signalled in language. He has not provided an account of how to read off the counterpart relation invoked by particular sentences. If context does not play exactly the role that Glick says it does then it might play some other role and the Glick theory might still be correct. For Glick there are, in fact, many g-relations corresponding to different relations of comparative similarity and different sentences invoke different g-relations. This allows Glick's theory to account for Edelberg-style cases by appealing to a shift in which g-relation is invoked by the respective sentences, even if it is not clear exactly how the g-relation being invoked shifts.

There are, it seems, two main approaches to accommodating Edelberg-style cases that we should to distinguish. One type of approach is what we might call *shifty* views, of which Glick's is an example. According to *shifty* views, the g-relation invoked shifts from sentence to sentence and that explains how one sentence gets to be true while the other is false. Another sort of view, what we might call *non-symmetric* views, says that in Edelberg cases the g-relation in question is not symmetric and that this explains the difference in truth values between the relevant sentences. For instance, we might say that at (f) Nob's belief has the same focus as Hob's but not vice versa. When I present my theory of intentional identity I will take the shifty option and *not* the non-symmetric option.

### 3.16. Conclusion

In the process of outlining the main alternatives to transparent theories in the literature, I have made a series of complaints. These complaints apply to different theories but all the theories I have discussed are subject to at least one of the complaints. The theories that look as though they are in the best shape are the counterpart theories, although they are still subject to the incompleteness objection. In the next chapter I will present a novel theory of intentional identity. Predictably, I will argue that none of the complaints that apply to the theories discussed in this chapter can be

levelled against the novel theory and that, for this and other reasons, the novel theory I propose ought to be preferred over the theories considered up to this point.

## CHAPTER 4: A NEW THEORY OF INTENTIONAL IDENTITY

### 4.1. The triangulation theory

In this chapter I propose a new theory of intentional identity, *the triangulation theory*, and argue that it has considerable advantages over its rivals. First, I illustrate the basics of the theory and illustrate how it works by showing how it delivers the verdict that (1) may be true in scenario (a) and by pointing out some of its most important features. Next, I argue that the triangulation theory has considerable advantages over its principal rivals. Finally, I extend and qualify the proposal and respond to some possible objections.

The central idea behind triangulation theory is that, when looking for conditions on intentional identity, we ought to look to the details of how agents take the putative target of their attitudes to be. As Geach points out, there are cases in which the target is missing but in which attitudes have a common subject matter. But what is *not* missing in such cases are the agent's *beliefs* concerning the putative target: its place in the world and what it would take for attitudes to be about it. Of course, if the attitudes are empty, many of these beliefs will be false, but the idea is that we can get a grip on intentional identity by having a detailed look at what agents believe about the putative targets of their beliefs.

I should remind the reader of the way that I am using the phrase 'putative target'. When I talk of 'putative targets' of attitudes, I mean the thing that an agent *thinks* the attitude is about. My talk of putative targets is somewhat rough and ready. When I talk about the targets of beliefs outside of intentional contexts, the use of the phrase 'the putative target' should be taken with a grain of salt. Again, it is one thing to think that there is an evil witch, it is quite another for there to *be* a witch that your attitude is about. See section 4.7.1 below for more details.

King (1993, 65) and Dennett (1968, 336-338, 341) adopt approaches which are somewhat similar to the triangulation theory in flavour. For King and Dennett intentional identity is grounded in the details of how the agents take the putative target to be. As we saw in chapter two, King's view and Dennett's view both run into apparently insuperable problems. The triangulation theory is my attempt to better implement a similar sort of approach.

For ease of reference I will repeat (a) and (1).



- (a) Hob and Nob live in the same village and read the same newspaper. The newspaper reports that a witch has been terrorising the village. Hob believes that a witch (the one mentioned in the newspaper) has blighted Bob's mare and Nob believes that the witch mentioned in the newspaper killed Cob's sow. Witches do not exist.
- (1) Hob thinks a witch has blighted Bob's mare, and Nob believes she (the same witch) killed Cob's sow.

Let us start by considering Nob and his beliefs in scenario (a) (my reasons for paying attention to Nob's beliefs first will become clear below). In scenario (a), Nob believes that there is a witch that has a certain place in the world: he believes that she is a cause of the newspaper article he read, and that some of his (Nob's) beliefs are about that witch. Nob is also disposed to judge that some other agents have beliefs about that witch. For instance, he judges that the author of the newspaper article has beliefs about the witch. He is also disposed to judge that certain other beliefs are *not* about the witch. He judges that the beliefs of agents who live alone in a cave on the other side of the world and who have never heard of the witch mentioned in the newspaper are not about the witch. Nob's judgements about these matters are not made arbitrarily. This plausibly reflects his beliefs about what, in general, it takes for beliefs to be about the putative target of his own belief. Nob believes that his belief came to be about the witch by being formed in response to his believing the reports made in the newspaper article. He also presumably believes that others may do the same, and thereby come to have beliefs about the same witch. That is, Nob presumably believes something like the following:

For any token belief B, if **B concerns a witch and was formed in response to believing newspaper article N**, then B is about the same thing as Nob's belief.

Call the part in bold a *triangulation condition*. By believing that, Nob *lays out* the part in bold as a triangulation condition for the putative target of his beliefs, in this case, the witch. Let us say that a belief B *satisfies* this triangulation condition if it concerns a witch and was formed in response to believing newspaper article N.

Subjects are likely to have many beliefs of this form, corresponding to the ways they think a belief might come to be about the putative target, given their beliefs about the target's place in the world.

I should note that many of these beliefs will be implicit. It is not required that any of these beliefs explicitly occur to the agent. These beliefs will be reflected in the agent's behavioural dispositions including their dispositions to judge, when confronted by a belief, whether that belief is about the witch. A good way to get an intuitive grasp on this idea is to imagine what would happen if Hob and Nob met up and had a discussion concerning the witch. On hearing about when and where Hob heard about the witch and the causal relationship that stands between his own belief and Hob's belief, Nob would be disposed to judge, on the basis of this information, whether or not the beliefs in question have the same putative target. The relevant beliefs will have the following general structure:

For any token belief B, if **B is such and such**, then belief B is about the same thing as belief A.

'A' stands for a particular token belief (in the original case A was Nob's belief concerning the witch). By believing a conditional of this form (and having another related belief, see the next paragraph), the subject *lays out* a triangulation condition 'belief B is such and such' for the target of A. A given belief *satisfies* a triangulation condition of the form 'belief B is such and such' when that belief is such and such. Some of these conditionals will concern particular objects or events; the example condition described above concerns a particular newspaper article. However, these conditionals can also be extremely general, concerning the target's place in the causal order of things broadly understood.

Note that we cannot extract a triangulation condition from *every* belief an agent has of this form. For reasons that will become clear later (in section 4.7.7.), we must insist that for an agent to lay down a triangulation condition for the target of an attitude by believing a conditional of that form, they must also believe that *if* a belief satisfies the condition in the antecedent of the conditional, then the belief in question is about that putative target *in virtue of* its satisfying the condition in the antecedent. This lines up with the characterisation of the source of triangulation conditions I gave above; the conditions agents lay out for the target are based on their beliefs about *what it takes* for an attitude to be about the target.

With the triangulation condition machinery in place we are in a position to state what, according to the triangulation theory, makes it the case that Nob's belief is g-related to Hob's in cases when

(1) is true. Nob's belief is g-related to Hob's belief, in the relevant sense, if, and only if, Hob's belief satisfies at least one of the triangulation conditions Nob lays out for the putative target of Nob's belief.

In scenario (a), Hob's belief does satisfy at least one of these triangulation conditions; it satisfies the example triangulation condition mentioned above. By believing that for any belief B, if **B concerns a witch and was formed in response to believing newspaper article N**, then B is about the same thing as his (Nob's) belief in virtue of being formed in response to believing newspaper article N, Nob lays out the triangulation condition 'belief B concerns a witch and was formed in response to believing newspaper article N' for the target of his belief. Hob's belief satisfies this condition. According to the triangulation theory, the fact that at least one of Nob's conditions is satisfied by Hob's belief explains why Nob's belief *is* g-related to Hob's belief in scenario (a), and, in turn, why (1) may be true in scenario (a).

There is a wrinkle that I have glossed over until now. Each g-relation corresponds to a particular set of triangulation conditions, laid out by a particular agent. G-relations are relative to which subject's triangulation conditions are in play. So we need a way of signalling whose triangulation conditions are in play when we are evaluating intentional identity in a particular case. Let us say that for A to be g-related<sup>S</sup> to B is for B to satisfy at least one of the triangulation conditions that a subject S lays out for the target of A. The superscript indicates whose triangulation conditions are in play.

Which agent's triangulation conditions are invoked varies from sentence to sentence. When evaluating (1) we focused on Nob's triangulation conditions. This is, in fact, what we did above when we ran through the details of how (1) gets to be true at (a); we talked as if being g-related in the sense relevant for the truth of (1) had to do with Nob's triangulation conditions. The idea is that, when we attempt to evaluate sentences like (1), we implicitly select some agent's standards for sameness of subject matter. We will come across more evidence for this claim in section 4.6, when I discuss how the triangulation theory helps us to handle the Edelberg-style cases that I discussed in chapter three. But why does (1) invoke Nob's triangulation conditions instead of Hob's, for instance? I propose that the location of the pronoun gives us some guidance as to whose triangulation conditions are invoked by a given sentence; (1) invokes Nob's triangulation conditions at least partly in virtue of the anaphoric pronoun 'she', occurring within an ascription of an attitude

to Nob. The rough idea is that we need to know who “she” is and we do that by checking to see who, among the putative targets of Hob’s beliefs, lines up with putative target of Nob’s belief *according to Nob*. This does not, of course, constitute a complete theory of how sentences invoke particular sets of triangulation conditions. I want to allow that *sometimes* the triangulation conditions laid out by the reporter (for example, the agent uttering a sentence like (1)) matter, instead of the triangulation conditions of one of the agents to which attitudes are being ascribed. Or perhaps it is the conditions of the evaluator of the sentence. My goal in this chapter is to present a theory of the relation of having a common focus that stands between intentional attitudes. My goal is not to present an account of how this relation is picked out in language. The issue of the semantics of sentences like (1) is delicate and complex, and no doubt deserves an in depth treatment, but that discussion is beyond the scope of the present project.<sup>16</sup>

We are now in a position to state the general schema that is the centrepiece of the triangulation theory:

(T) For all token beliefs A and B, A is g-related<sup>S</sup> to B if, and only if, B satisfies at least one of the triangulation conditions S lays out for the putative target of A.<sup>17</sup>

Let us run through the evaluation of (1) in scenario (a) in the interests of illustrating how (T) works and how to plug in the appropriate beliefs and subject. When evaluating (1) we need to check whether Nob’s belief is g-related<sup>Nob</sup> to Hob’s belief. Making the appropriate substitutions, we get: Nob’s belief is g-related<sup>Nob</sup> to Hob’s belief if, and only if, Hob’s belief satisfies at least one of the triangulation conditions Nob lays out for the target of Nob’s belief. This is the triangulation theoretic a test for intentional identity in the original case.

---

<sup>16</sup> For a discussion of some of the central complexities surrounding the semantics of anaphoric pronouns see Heim (1982, 1990).

<sup>17</sup> This schema is adapted to handle beliefs with only one relevant target. If the triangulation theory is going to be developed so as to handle beliefs with more than one relevant target, there will be an even more general variant on (T) which will be relativized to a particular target of A. I will generalize the theory in section 4.7.1.

As discussed above, the right-hand side of this conditional is true at (a) so Nob's belief *is* g-related<sup>Nob</sup> to Hob's belief (there is intentional identity between their attitudes) and (1) gets to be true at (a).

Cases in which two or more of a single agent's beliefs are g-related are to be handled in a similar way. Nothing in the triangulation theory requires that the beliefs that we feed into (T) be beliefs of distinct agents, for more information on this sort of case see section 4.7.8.

Note that, for the triangulation theorist, beliefs can be g-related<sup>S</sup> without S believing that they are g-related. What is required is that at least one of the relevant triangulation conditions is, in fact, met by the belief; satisfaction of one of the relevant set of conditions is what matters, not beliefs about satisfaction. For instance, (1) can be true, when Nob has no beliefs whatsoever about Hob, Hob's beliefs or Bob's mare. This feature of the view will be crucial in chapters five and six when I discuss how this theory can form the centrepiece of a general package of views concerning the content of intentional states. It also means that the triangulation theorist does not fall into the trap that King's theory and one kind of lazy theory do by requiring that Nob have beliefs concerning Hob for there to be intentional identity between their beliefs.

It is also worth mentioning that the triangulation theory applies to sentences that invoke more than one g-relation. For instance, consider (12).

(12) Hob thinks a witch blighted Bob's mare, Nob and Rob believe that she (the same witch) killed Cob's sow.

(12) invokes the triangulation conditions of both Nob and Rob. The structure of (12) means that we have to run (T) twice, once to check if Nob's belief is g-related<sup>Nob</sup> to Hob's belief and once to check if Rob's belief is g-related<sup>Rob</sup> to Hob's belief. Nothing in the triangulation theory implies that at most one set of triangulation conditions can be relevant when evaluating a given sentence.

#### 4.2. Intentional distinctness

The triangulation theory delivers the correct verdict that (1) is false at cases like (c).

(c) Hob, a butcher in Arkansas, believes that a witch lurks under his bed and she has blighted the mare of his neighbour Bob. Nob, a baker in Ukraine, believes that a witch lurks under

his (Nob's) bed and that she killed the sow of Nob's neighbour Cob. Hob and Nob have never met and there has been no relevant interaction between their social circles. Witches do not exist.

Hob's belief plausibly does not meet any of Nob's triangulation conditions, in scenario (c). In scenario (c) there is no newspaper that they both read, and there is no single act of livestock malingering that is a common cause of their respective beliefs. Since Hob's belief does not satisfy any of Nob's triangulation conditions in scenario (c), by (T), Nob's belief is not g-related<sup>Nob</sup> to Hob's, so (1) is false in scenario (c).

#### 4.3. Disagreement

The triangulation theorist can deliver the correct verdict that there may be intentional identity in cases in which the agents have conflicting beliefs about the target. For example, (1) may be true in scenario (b).

- (b) Hob and Nob live in the same village and read the same newspaper. The newspaper reports that a witch has been terrorising the village. Hob believes that the witch mentioned in the newspaper blighted Bob's mare. Nob thinks the witch mentioned in the newspaper did not blight Bob's mare but did kill Cob's sow. Witches do not exist.

In scenario (b), just as in scenario (a), one of Nob's triangulation conditions are met as regards Hob's belief. Of course, on learning about Hob's belief, Nob is disposed to judge that one or other of the two beliefs must be false, but this does not threaten the truth of (1) in scenario (b). A belief being *compatible* with Nob's belief is, according to Nob, not a necessary condition for that belief and his (Nob's) own belief to be about the same target. This is as it should be; Nob believes that witches are the kind of thing that people can have conflicting beliefs about.

#### 4.4. Flexibility

The triangulation theory is desirably flexible. The triangulation conditions discussed above have been predominantly causal. These triangulation conditions suit how Nob takes the target of his belief to be. But the triangulation theory is adaptable to cases in which the believer does not believe that the target of her belief is *that* kind of object.

A subject may think that the target of a belief is such that it can be picked out without any subject having had any causal contact with the target. For instance, if the target of Nob's belief were a mathematical object, then he would likely not believe that for a belief to be about that object the belief must be causally related to the object. He is more likely to believe that having a belief about that object is a matter of the believer describing the target object in the right way, or allowing it to play a certain role in the believer's mathematical reasoning. Triangulation is sometimes made conditional on causal links and sometimes not.

Agents can lay out triangulation conditions for the target that yield extremely generous g-relations. Consider Spinoza's beliefs about God. Suppose there is no God and Spinoza's beliefs about God are empty. Suppose that Spinoza also believes that God is literally identical to everything in the world. For a belief to be about God, according to Spinoza, it just has to be about *something*. Accordingly, the set of triangulation conditions Spinoza lays out for the target make for an extremely generous g-relation.

Near the other end of the spectrum, triangulation conditions can be extremely restrictive. Suppose Nob thinks he is the chosen one and that an angel whispers instructions in his ear. He is hallucinating, there is no such angel. He may think that almost nobody else could have beliefs about that angel. In this case he is likely to judge that very few other beliefs, if any, are about the angel. This is an example of an extremely restrictive set of triangulation conditions for the target of a belief.

Agents *can* have extremely peculiar views about the target entity and about how beliefs get to be about what they are about and may, consequently, lay out correspondingly peculiar triangulation conditions. When agents lay out unruly triangulation conditions for the putative targets of their beliefs, their beliefs about the target do not appear to be different in kind from the other beliefs we have considered thus far. I will discuss cases in which agents lay out peculiar triangulation conditions more fully in sections 4.7.9 and 4.8.1.

Agents who lay out unruly triangulation conditions are likely to be criticisable in many respects. For example, they might have trouble co-ordinating their beliefs with others and communicating about the target, especially if others lay out wildly different triangulation conditions for the target. There is nothing in the triangulation theory that disallows agents from laying out unruly triangu-

lation conditions. The triangulation theory does not yield answers about which triangulation conditions agents ought to lay out. It only concerns how g-relatedness depends on the triangulation conditions that subjects do, in fact, lay out. When we are dealing with an agent with peculiar views about how things are we should expect verdicts about g-relations to be correspondingly peculiar.

There is, however, one sort of case involving an agent laying down aberrant triangulation conditions that I ought to discuss explicitly and give special treatment to. An agent could lay down triangulation conditions such that according to her beliefs, belief A is *not* about what it is about. For instance, an agent could believe that if a belief B is *not* about what A is about then B is about what A is about. The triangulation theorist ought to say that in such a case, his own belief, A, fails to be about anything at all. In such a case, A cannot be g-related to anything, not even to itself. If a set of triangulation conditions would yield a g-relation that is non-reflexive, those triangulation conditions do not yield a g-relation at all.

#### 4.5. No appeal to exotica

Note that the triangulation theory does not involve an appeal to any merely possible objects, mythical objects or non-existent objects. According to the triangulation theorist, the g-relation does not depend in any way on the objects, if there are any, which the relevant beliefs are about. The triangulation theory should not be understood as a way of providing conditions on the identity or distinctness of intentional objects. It is a theory of when beliefs are g-related; it is a theory about when and why beliefs have a common focus, in the sense Geach discussed.

The label ‘intentional identity’ is somewhat misleading in that it suggests that intentional identity ought to be understood as bound up with the identity conditions of the objects of intentional attitudes. But this is just one way of understanding intentional identity that the triangulation theorist need not adopt.

This is not to say that the triangulation theorist ought to claim that there are no such exotic objects; it may be that in light of some other considerations we will be driven to countenance them. All the triangulation theorist will claim is that we are not required to countenance or make reference to exotic objects in our explanation of intentional identity.



#### 4.6. The advantages of the triangulation theory

The triangulation theory is not subject to the complaints mentioned in the previous chapters. The triangulation theory is at least a good deal *more* complete than the theories of Salmon, Parsons, Geach, Glick and van Rooy. As we saw in sections 4.1 to 4.3 inclusive, the triangulation theorist can provide conditions on intentional identity in particular cases. G-relations vary with triangulation conditions and these vary across agents and from scenario to scenario, but the triangulation theory will deliver an explanation of *how* the conditions vary from case to case. It will a schema along the lines of (T) that, when the relevant beliefs and conditions are plugged in, will deliver verdicts on intentional identity in particular cases.

For the triangulation theorist the question of *which* causal links matter to intentional identity has a straightforward answer: the causal properties that matter are those which are mentioned in the relevant triangulation conditions. When evaluating (1) in scenario (a) we considered Nob's triangulation conditions. Based on his beliefs about the putative target of his own belief and its place in the world, he lays down triangulation conditions that concern the causal history of the beliefs. For this reason, the fact that Hob's belief and Nob's belief are g-related in scenario (a) is explained by their having a certain kind of causal history. Causal relations make a difference to intentional identity *because* the triangulation conditions in play concern causal relations of that kind. Also, it is these causal relations that matter rather than some others, because the relevant triangulation conditions concern those causal relations rather than others.

The triangulation theory easily handles cases in which there is intentional identity but no causal link between the relevant beliefs. Some triangulation conditions concern causal relations, others do not. In cases in which there is intentional identity but no causal link, the triangulation theory can still deliver the right verdict. Take, for instance, the Leibniz/Newton case discussed in section 3.10., repeated here for ease of reference. Suppose Leibniz and Newton, quite independently of one another, both come up with the idea of a mathematical function *integration* (though they use different words to refer to it) that plays almost exactly the same role in their respective mathematical reasoning. Suppose further that they are mathematical Platonists; they think that mathematical objects are abstracta that exist in Plato's heaven but they are wrong, there are no abstract mathematical objects. In this case there is no causal link of the appropriate kind between the two beliefs since they came up with the idea independently. Yet it seems as if their beliefs may be g-related;

their beliefs have a common focus in being directed at a particular function.

The triangulation theorist can explain how these beliefs can be g-related in the absence of a relevant causal chain. Since in this case Leibniz and Newton are both mathematical Platonists, they are likely to think that mathematical objects such as functions do not causally interact with concrete objects. For this reason they are *unlikely* to believe that, in order for beliefs to be about such things they must be causally downstream of the mathematical objects. Their beliefs concerning the putative target's place in the world are reflected in the triangulation conditions they lay out for the putative target. Leibniz might, for instance, believe that for any belief B, if B plays a certain role R in the believer's mathematical reasoning, then B is about the same thing as A (where A is one of Leibniz's beliefs putatively about the integration function). Newton's belief might, in fact, satisfy one such condition that we can extract from this belief and we would thereby reach the correct verdict that the relevant beliefs are g-related.

These features of the triangulation theory concerning causal links can be understood by means of drawing a parallel to *causal descriptivist* theories of reference. Causal description theory was developed in the face of arguments presented by Donnellan (1970), Kripke (1980), Putnam (1975) and others against the traditional theory of descriptions defended by Russell (1905, 1910-1911). According to Russell, ordinary proper names are usually descriptions in disguise. The idea is that these names refer like descriptions refer, by mentioning some properties that the thing named is supposed to have. The upshot of the Donnellan-Kripke-Putnam arguments is supposed to be that most ordinary proper names do not refer in this way; no description will do the job. The alternative picture suggested was that names usually refer by means of a causal chain of communication leading back from token use of the name to an initial dubbing event by which a name was given to an object, at least in the good case.

The casual descriptivists like Loar (1976), Kroon (1987, 2009), Lewis (1984, 1997) and Jackson (1998, 2010) say, to borrow a phrase from Lewis, that descriptivists have learnt enough from their opponents to be immune to their attacks. The idea is that arguments presented against the description theory relied on an overly simplistic understanding of the descriptions that are supposed to capture the semantic value of a name. In particular they relied on a kind of 'famous deed' descriptivism according to which the description associated with the name 'Aristotle' is something like 'the teacher of Alexander the Great'. Once we move away from this simplistic model and allow

that the descriptions in question might be more complicated than we originally thought, the descriptivist avoids the problem presented by Donnellan, Kripke and Putnam. What the Donnellan-Kripke-Putnam arguments really show is that certain simplistic descriptions cannot be what captures the semantic content of some names. In a slogan, borrowed from Lewis, when causal theory works a causal descriptivism works too.<sup>18</sup> , the descriptivist will take any causal condition on reference presented by causal theorists and build those conditions into the relevant causal description. In this way, the causal description theorist can deliver all the same verdicts as the pure causal theorist by putting the appropriate causal properties into the relevant description. The causal descriptivist theory is *flexible*, some of the relevant descriptions involve causal properties, and others do not. In the same way, the triangulation theory can mimic the behaviour of the theories based on the causal link thesis while, at the same time, handling the non-causal cases that those theories struggle with.

Another appealing feature of the triangulation theory is that, if it works, it will allow us to tell a uniform story about subject matter individuation across empty and non-empty cases. As mentioned above, the triangulation theory does not require there to actually be an object that the relevant attitudes are about. For the triangulation theorist, the presence or absence of the object does not, in general, make a difference to questions of subject matter individuation. All else being equal, we ought to prefer theories of subject matter that treat empty and non-empty cases uniformly. In chapter two we saw one attempt to maintain uniformity by treating apparently empty beliefs as being about objects after all. As we saw, that sort of approach ran into serious problems. The triangulation theory promises to treat the two sorts of cases the same. The theory was developed so as to handle the hard cases but, if uniformity is a virtue and the theory works in non-empty cases, there is reason to apply it to the non-empty cases as well.

As for Edelberg-style cases, the triangulation theorist is in a position to deliver the verdict that (1) is true in scenario (f) while (7) is false. I will repeat (f), (1) and (7) here for ease of reference.

- (f) Hob and Nob are partners in a witch-hunting business. They investigate the suspicious illness of Bob's mare. They both conclude that Bob's mare was blighted by a witch who was working alone. The next day they investigate the suspicious death of Cob's sow. At

---

<sup>18</sup> Kripke (1980, 88 n.38) concedes this.

this stage a disagreement arises between Hob and Nob. Nob thinks one witch both blighted Bob's mare *and* killed Cob's sow. Hob thinks one witch blighted Bob's mare and a different witch killed Cob's sow. Witches do not exist.

- (1) Hob thinks a witch has blighted Bob's mare, and Nob believes she (the same witch) killed Cob's sow.
- (7) Nob thinks a witch has killed Cob's sow, and Hob believes she (the same witch) blighted Bob's mare.

Recall that, for the triangulation theorist, there are many g-relations that correspond to different sets of triangulation conditions laid out by different agents and different sentences invoke different subjects' triangulation conditions. In scenario (f), Hob and Nob plausibly lay out different triangulation conditions for the target of their respective beliefs; they have different beliefs about which witches are identical to which.

The triangulation theorist thus leaves room for (1) and (7) to differ in truth value in scenarios like (f). Plausibly, (1) invokes Nob's triangulation conditions and (7) invokes Hob's triangulation conditions. In (1) the pronoun 'she' occurs in an ascription of an attitude to Nob, whereas in (7) it occurs within an ascription of an attitude to Hob. If the conjecture I made above is right, this fact would explain why (1) invokes Nob's triangulation conditions and (7) invokes Hob's. Understanding the invocation of triangulation conditions in this way also matches intuitive judgements about the truth values of (1) and (7) in scenarios like (f).<sup>19</sup> If the two sentences invoke different triangulation conditions, and Hob and Nob lay out different triangulation conditions for the target their respective beliefs, then (1) can be true while (7) is false. Hob's belief meets at least one of the triangulation conditions that Nob lays out for the target of his (Nob's) belief in scenario (f), so (1) gets to be true at (f). But Nob's belief does not satisfy any of the triangulation conditions Hob lays

---

<sup>19</sup> Even if I am wrong about *how* different agent's triangulation conditions are invoked in language, the triangulation theory will still have this advantage. Since there are many g-relations, if the triangulation theorist is right, the theory leaves room for some sort of shift between the g-relations invoked by (1) and (7), even if it is unclear exactly *how* which g-relation is invoked shifts from sentence to sentence.

out for the target of his (Hob's) belief, so (7) comes out false at (f). In this way the triangulation theorist can deliver the correct verdicts in Edelberg-style cases.<sup>20</sup>

Note that this way of approaching Edelberg-style cases is an improvement on Glick's strategy, couched within his counterpart theory, which I discussed near the end of the previous chapter. The triangulation theorist does not make intentional identity beholden to the content of the first conjunct of some intentional identity sentences. Consider (10), repeated here for ease of reference.

(10) Hob thinks a witch blighted Bob's mare, and Nob believes she (the same witch) did not blight Bob's mare.

Remember that Glick's strategy fails when confronted with sentences like (10). It did not lead us to the correct verdict that (10) is true at (b), the original case of disagreement. Glick's strategy is also unable to deliver the verdict that (1) gets to be true in scenario (b). If we employ Glick's strategy, the kind of disagreement that is present in (b) unduly threatens the truth of sentences like (1).

As we saw in section 4.3 above, the triangulation theorist can explain why (1) gets to be true at (b), repeated here for ease of reference.

(b) Hob and Nob live in the same village and read the same newspaper. The newspaper reports that a witch has been terrorising the village. Hob believes that the witch mentioned in the newspaper blighted Bob's mare. Nob thinks the witch mentioned in the newspaper did not blight Bob's mare but did kill Cob's sow. Witches do not exist.

As we saw in section 3.14, Glick's approach does not yield this result. What is more, the transparent theorist can also deliver the correct verdict that (10) is true at (b). The content of the belief mentioned in (10)'s first conjunct need not necessarily make a difference to the relevant kind of g-relatedness. If (10) invokes Nob's triangulation conditions for the relevant target, and Nob lays

---

<sup>20</sup> This handling of Edelberg-style cases does *not* involve an asymmetric g-relation, in the terminology of the last chapter, the triangulation theorist will appeal to a shifty strategy, rather than a non-symmetric strategy for dealing with Edelberg-style cases.

out triangulation conditions that allow for intentional identity in spite of the disagreement in question, then (10) gets to be true in scenario (b).

#### 4.7. Extensions and qualifications

I have now presented the basics of the theory and argued that the theory has some considerable advantages over its rivals. Next, I will clarify and extend the account somewhat.

##### 4.7.1. Attitudes with more than one focus

It is time to extend the theory so it can handle attitudes with more than one focus. Paradigm cases of attitudes with more than one focus are those concerning relations. For instance, I might believe that Jane is smarter than Jill or desire that I be richer than Sally. Consider scenario (g).

- (g) Hob and Nob live in the same village. They are discussing the various witches they believe have been causing trouble around the village. They both believe that there are three witches who they call Esmerelda, Harriet and Meg. Hob believes that Harriet is taller than Esmerelda. Nob believes that Harriet is taller than Meg. Witches do not exist.

Now, are the beliefs mentioned in (g) g-related? Yes and no. They are both directed at Harriet and, in this sense, they have a common focus. On the other hand, one has Esmerelda and not Meg as a putative target, the other has Meg and not Esmerelda as a putative target.

How can the triangulation theory be adjusted so as to capture the details of this sort of case? I propose that another index be added to (T) that relativises g-relations to putative targets, as well as to an agent's triangulation conditions.

- (T2) For all token beliefs A and B, A is g-related<sub>O</sub><sup>S</sup> to B if, and only if, B satisfies at least one of the triangulation conditions S lays out for O.

The O in (T2) stands for the conception that S employs to pick out a particular putative target of A. O represents S's conception of a putative target of A. Let us see how this works when applied to scenario (g). Remember that triangulation conditions are laid out for particular targets that the attitudes in question are *supposed* to be about. An agent will often lay out different triangulation conditions for different targets and there is nothing stopping an attitude being aimed at more than

one object. The fact that they believe the attitude has more than one focus will be reflected in their second-order beliefs concerning the various putative targets of the relevant belief.

To evaluate whether Hob's belief and Nob's belief both focus on the putative target Harriet, we make substitutions into (T2) and arrive at the following bi-conditional: Nob's belief is g-related $_{\text{Harriet}}^{\text{Nob}}$  to Hob's belief if, and only if, B satisfies the triangulation conditions Nob lays out *for Harriet*. In scenario (g) the right hand side is satisfied so that means that the relevant beliefs are plausibly g-related in that sense, that is, they are g-related $_{\text{Harriet}}^{\text{Nob}}$ .

What about the sense in which they are not g-related? We see that if we substitute 'Meg' for 'Harriet' in the conditional the right hand side is plausibly *not* satisfied since Nob believes that Meg is distinct from Harriet. So the attitudes mentioned in (g) need not be g-related $_{\text{Meg}}^{\text{Nob}}$ .

In spite of the complexity (T2), the idea behind this relativisation to a putative target is compelling. Triangulation conditions are laid out for putative *targets of attitudes* rather than simply for attitudes. Attitudes often have a fairly rich structure and complex attitudes are often about more things than simple. This complexity ought to be reflected in our theories concerning the subject matter of those attitudes.

Two points are important to note. Firstly, the inclusion of O into the schema does not mean that for either side of the relevant bi-conditional to be satisfied there needs to be an object that O is a conception *of*. That is to say, this relativisation does not commit the triangulation theorist to intentional objects. The person to whose triangulation conditions the relevant g-relation is relativised will have a view about what the object corresponding to O is like. That agent might be quite wrong about what there is, for instance, it might be that there is no such object as O. This is, in fact, the case in scenario (g). Nob is wrong about there being an object that corresponds to his conception of Harriet. O represents not an object O but *O as conceived by S*.

Secondly, these 'conceptions' of objects that I appeal to need not have many of the interesting properties often associated with *concepts*. For instance, conceptions need not be shared across agents or by an agent across time. They are private conceptualisations of putative targets of beliefs on the part of the agent.

#### 4.7.2. Dependent triangulation conditions

There is a certain class of triangulation conditions that I ought to address directly.<sup>21</sup> Agents often lay out what I will call *dependent* triangulation conditions for a target entity. For instance, in scenario (a), Nob presumably believes the following conditional.

For any belief B, if B is about the witch mentioned in the newspaper, then B is about the same thing as A is.

By believing this conditional, Nob lays out ‘belief B is about the witch mentioned in the newspaper’ as a triangulation condition for the target of A. Does Hob’s belief satisfy this condition at (a)? The question is tricky, since the condition concerns what B is *about*.

Another kind of dependent triangulation condition is laid out by agents who believe the following conditional.

For any belief B, if B is about the same as A is, then B is about the same thing as A is.

The triangulation condition laid out by agents who believe this sort of conditional is ‘belief B is about the same thing as A is’. Supposing Nob lays out this condition for the target of his belief, is it satisfied by Hob’s belief at scenario (a)?

What the triangulation theorist ought to say about dependent conditions is that they are satisfied by belief B if, and only if, some *other* triangulation condition the agent lays out for the putative target is met by B. In all other cases the dependent conditions are not satisfied. That is, the satisfaction of a dependent triangulation condition is *dependent* on the satisfaction of some other condition laid out for the target. Hob’s belief satisfies some other triangulation condition laid out for the target of A, for example, the one that mentions being formed in response to the newspaper article. Since the dependent conditions concern subject matter, we need to evaluate the relevant attitudes and what they are about, by evaluating the *other* conditions on attitudes focusing on that putative target, before we determine whether or not the dependent condition is satisfied by a given belief.

---

<sup>21</sup> I am grateful to Racheal Briggs and Robbie Williams for impressing upon me the importance of handling these sorts of triangulation conditions.



### 4.7.3. Other attitudes

The triangulation theory can be naturally extended to handle cases involving other intentional attitudes. Agents have beliefs about what it takes for a belief to be about a particular object, given that object's place in the world. They also presumably have beliefs about what it takes for a desire to be about that object, or for a fear to be about that object, and so on. The triangulation conditions discussed above concerned beliefs. But presumably, agents also have beliefs of a similar form concerning other kinds of attitudes. That is to say they believe conditionals of the following form

For any token attitude B, if B is such and such, then B is about the same thing as A is.

These beliefs guide the agent's judgements concerning what attitudes are about. Since these beliefs are plausibly in place, we can extend the triangulation theory to handle other sorts of attitudes, in a straightforward way. We need only move from (T2) to (T3), by replacing the term 'belief' with the term 'attitude'.

(T3) For all token attitudes A and B, A is g-related<sub>O</sub><sup>S</sup> to B if, and only if, B satisfies at least one of the triangulation conditions S lays out for O.

### 4.7.4. Attitudes about kinds and properties

Attitudes concerning classes of things and properties can also have a common focus. Consider a scenario in which two scientists, Jack and Jill, are exchanging views about Phlogiston. The following sentence might be true.

(13) Jack thinks Phlogiston is a gas, Jill believes that it (the same stuff) was released by process P.

(13) ascribes intentional identity in that for (13) to be true Jack's belief and Jill's belief have to be g-related.

The triangulation theorist will handle this sort of case in roughly the same way it handles intentional identity between attitudes that are directed at particular objects. Agents lay out triangulation conditions for the target stuff in much the same way that they lay them out for particular targets. They will base these beliefs on their beliefs concerning the stuff in question and its place in the

world. The test for intentional identity will be exactly as before, except that the putative target of the beliefs will be a kind or a type of thing rather than a particular object.

A similar move can be made to handle attitudes that have a property as their focus. Consider a case in which two philosophers Locke and Key are arguing about the status of the colour red and its relation to objects. (14) may be true in such a case.

(14) Locke thinks redness is not inherent in objects, Key believes that it (the same property) is inherent in objects.

The triangulation theorist will want to handle this sort of case in the now familiar way. Agents will lay out triangulation conditions for the target property or properties and other attitudes will be g-related, in virtue of both focusing on the same property, just in case the invoked triangulation conditions are met by the relevant attitudes.

#### 4.7.5. Triangulating with more than one agent

So far I have considered cases involving just one attitude having to meet the relevant triangulation conditions. There is no reason why some cases might require us to run the triangulation theory's test for more than one belief.

(15) Hob and Rob both think a witch has blighted Bob's mare, Nob believes she (the same witch) killed Cob's sow

For (15) to be true Hob's belief must be g-related<sup>Nob</sup> to Nob's belief and Rob's belief must be g-related<sup>Nob</sup> to Nob's belief. This is an instance of having to run the triangulation condition test twice, once for each relevant attitude.

#### 4.7.6. Symmetric intentional identity ascriptions

Recall (2).

(2) Sal and Val fear the same werewolf.

When evaluating (1) and (7) above we took our guidance about which agent's triangulation con-

ditions where invoked by those sentences from the location of the pronoun. In (2) there is no pronoun. There are two obvious options concerning whose triangulation conditions are invoked here. The first is that it invokes *both* Sal's conditions and Val's conditions. On this reading, the sentence would require both that Sal's fear met Val's triangulation conditions, and that Val's fear met Sal's triangulation conditions. We simply need to run the test in both directions for the respective g-relations.

Another option is that the triangulation conditions invoked are those of the reporter, the utterer of (2), or the evaluator, the agent who is evaluating the sentence. If this is the right reading and we are taking the reporter's or the evaluator's standards for sameness then we need only run the test as usual, plugging in the appropriate attitudes and triangulation conditions.

#### 4.7.7. Ignorance of g-relatedness and mistaken beliefs about g-relatedness

Agents can be wrong about whether or not an attitude has a common focus as another attitude, *by that agent's own standards*. Consider a case in which there are two newspaper articles that appear in the local paper, each claiming that there is a witch terrorising the village. Suppose on reading both articles, Nob comes to the belief that one of the newspaper articles (N1) is correct and the other (N2) is a pack of lies. He also lays down some triangulation conditions for the target of his newly formed witch beliefs. In particular, he believes that if a given belief was formed in response to N1 then it is about the putative target of his own belief. Hob only reads N2, which Nob thinks is a pack of lies.

Later that day, Hob and Nob meet in the pub and Hob tells Nob about his (Hob's) suspicions concerning the witch. He believes that the witch mentioned in the newspaper has blighted Bob's mare. Nob mistakenly believes that Hob has read both articles and has formed the belief he is expressing in response to reading and believing N1, much as Nob himself did. For this reason, combined with his belief in the conditional mentioned above, he forms the belief that his belief and Hob's belief are about the same target. In this case, however, he is wrong about the causal history of Hob's belief. In fact, Hob's belief was not formed in response to N1 but in response to N2.

Hob's belief does not meet the triangulation conditions Nob lays out for the putative target of his

belief but Nob believes that it does. So long as Nob's triangulation conditions are invoked, (1) may very well be false at this scenario. Even though Nob *believes* that their beliefs have a common focus by his standards, he may be wrong about that.

Another way that an agent can be wrong about which beliefs have a common focus by their standards is a case like the one above except in reverse. Nob might believe that the beliefs do not have a common focus while *by his own standards* they do. The lesson should be clear. If the triangulation theory is correct, an agent's beliefs about what it takes for an attitude to be about a putative target do not dictate which attitudes are g-related, they only dictate the conditions under which they are g-related. The agent may be mistaken about whether these conditions are met by a given attitude.

Cases like these also reveal the need to constrain the extraction of triangulation conditions from an agent's beliefs in the way described in section 4.1. We have to insist that for an agent to lay down a triangulation condition for the target of an attitude by believing a conditional of the following form,

For any token belief B, if **B is such and such**, then belief B is about the same thing as belief A,

they must also believe that *if* a belief satisfies the condition in the antecedent of the conditional, then the belief in question is about that putative target *in virtue of* its satisfying the condition in the antecedent.

If we do not put this constraint on triangulation condition extraction, Nob's mistaken belief that they had the same target in mind threaten to undermine the claim that he is mistaken at all about this in the case described. Nob would presumably believe that *if* Hob's belief is about whatever Hob and Nob were talking about at the pub, then it is about what his (Nob's) belief is about. But if we take this to involve laying out a triangulation condition for the target, we end up being committed to saying that Nob is not *really* mistaken about his and Hob's beliefs having a common focus.

The 'in virtue of' constraint on an agent laying down a triangulation condition will get me out of this problem. Although Nob plausibly does believe that conditional, he presumably does not believe that Hob's belief is about the witch *in virtue of* being the subject of their pub-talk. For this

reason the claim that Nob can be mistaken about g-relations in the way described above is not threatened in this way.

#### 4.7.8. Intra-personal intentional identity

Another illustrative class of cases involve intra-personal intentional identity, when two or more of one agent's beliefs have a common focus. Suppose Hob believes that the witch mentioned in the newspaper is evil and that the witch mentioned in the newspaper is cunning. If we plug Hob's beliefs into T3 and make the appropriate substitutions we can test whether or not these two beliefs have a common focus in just the same way as T3 can be used as a test in cases in which the attitudes in question are the attitudes of more than one agent.

Just as in the inter-personal case, for the attitudes to have a common focus it is not required that the agent believes that they have a common focus. Suppose Hob, after reading about the witch in the newspaper, forms the belief that the witch mentioned in the newspaper is evil. Later that day, Hob is at the pub gossiping with Rob. Rob tells Hob about his (Rob's) suspicions about there being a witch that is terrorising the village. Hob forms the belief that the witch that Rob has in mind is cunning. However, Rob does not let on whether or not he developed this suspicion in response to reading the same newspaper as Hob.

In such a case, Hob may wonder whether or not his two beliefs have a common focus, that is, he might wonder if the witch that Rob was talking about is the witch that was mentioned in the newspaper. This is a perfectly sensible thing for Hob to wonder. In this case Hob does not believe that the beliefs have a common focus but it may be that, nonetheless, these two attitudes have a common focus *by Hob's own lights*. That is to say, it may be that at least one of the triangulation conditions Hob lays out for the putative target of one of his beliefs is, in fact, satisfied by this other belief, even if he does not believe that this is the case. What is more, and this relates to the point made in the last section, he may also have beliefs that, by his own lights, are mistaken concerning whether or not two of his beliefs have a common focus. The general lesson is that according to the triangulation theory, when it comes to subject matter individuation, intrapersonal and interpersonal intentional identity cases are not necessarily to be treated differently.

#### 4.7.9. Strange triangulation conditions, the reporter and the evaluator

How does the truth of intentional identity claims interact with agents that lay out peculiar triangulation conditions? The triangulation theorist ought to say that the triangulation conditions *can* be invoked. Suppose Jane believes her dog is tired and that Spinoza (who believes that everything is literally identical to God) believes that God is wise. If we explicitly invoke Spinoza's very generous standards on sameness of subject matter so that it is his triangulation conditions that matter, then the following claim comes out true: Jane thinks her dog is tired, Spinoza thinks it, the same object, is wise. Note that if Jane's conditions are invoked and Jane is less generous about triangulation then things turn out differently. If Jane's conditions are invoked the following claim will likely come out false: Spinoza thinks God is wise, and Jane thinks it, the same object, is tired.

The way to think about these cases is that we can *force* peculiar triangulation conditions to be invoked but that this need not be common. When evaluating intentional identity claims that involve agents with peculiar views about the putative target or what it would take for an attitude to be about that target, we sometimes implicitly invoke either the triangulation conditions of the reporter, the person making the intentional identity claim, or the evaluator, the agent who is attempting to evaluate the claim.

#### 4.7.10. Sortal relativity

Intentional identity is sometimes *sortal relative*. To use an example due to Lewis (1993), suppose you hold out a copy of the Tuesday Age and a copy of the Wednesday Age and ask 'same or different?' If you mean same newspaper, I know what to say. If you mean same issue, I know what to say. But until I know which standard for sameness you are asking about, I don't know how to answer.

In a similar way, conditions on intentional identity will sometimes be sortal relative. Intuitively, two beliefs can be about the same newspaper *issue* without being about the same *copy* of that newspaper. It is one thing to say that the copy of the newspaper is a common focus of two or more attitudes, it is another to say that the newspaper issue is the common focus of two or more attitudes.

The triangulation theorist can easily make sense of this kind of sortal relativity. Agents will, it seems, lay out different triangulation conditions for an object *under different sortals*. Suppose I

have a belief about the newspaper in front of me. I believe that some attitudes are about this newspaper qua issue without being about this newspaper qua copy. For instance, I might believe that other readers who bought a different copy of the same issue would have beliefs about this newspaper qua issue but not about this newspaper qua copy. This suggests that I believe that the conditions on attitudes being about the same *issue* as my belief is are different from the conditions on attitudes being about the same *copy* as my belief is. These will be reflected in the triangulation conditions I lay out for the putative target of my belief under these two sortals, *issue* and *copy*.

In light of this sortal relativity we ought to refine T3 so that we end up with T4.

(T4) For all token attitudes A and B, A is g-related $_{OQ}^S$  to B if, and only if, B satisfies at least one of the triangulation conditions S lays out for O, under the sortal Q.

Often which sortal is in play will not matter and T3 will suffice. For the sake of simplicity I will often employ T3, suppressing the sortal relativity, unless sortal relativity is relevant to the particular case. Nonetheless, we should keep T4 in our back pocket.

It is also worth noting that agents often have beliefs about how the different sortals that an object falls under relate to each other and how this interacts with intentional identity. For instance, when I have a copy of a newspaper in front of me, I believe that if an attitude is about this newspaper qua copy it is also about this newspaper qua physical object. But I do not believe that if an attitude is about this newspaper qua issue, then it is about this newspaper qua physical object. An agent's conception of the putative target of their belief is often very rich and involves information about which sortals the target falls under and the relationship between those sortals.

#### 4.7.11. Indexing to an agent at a time

As an agent's conception of a putative target changes over time, the triangulation conditions they lay out for that putative target may also change. For this reason, the index that makes the g-relation relative to a particular agent's triangulation conditions that features in the schemas T through T4 should be understood as relativizing the g-relation to an agent *at a time*. The cases I consider in this thesis do not involve a change in the triangulation conditions an agent lays out for the putative target, so the g-relation being relative to a time will not make a difference.

#### 4.7.12. Toy examples and the overall strategy

In the interests of illustrating the mechanics of the theory I have been discussing more or less toy examples of triangulation conditions and the kind of general beliefs about attitudes from which I propose we can extract the triangulation conditions. I admit that the example conditions described above are likely to be far too simplistic. The triangulation conditions that people actually lay down are likely to be vastly more complicated. It may be that the beliefs that guide agents' judgements about subject matter are too complicated to be expressed precisely in manageable sentences that, say, a person could state precisely in their PhD thesis. But once the basic framework is clear, an increase in the complexity of the triangulation conditions in question will not, as far as I can see, threaten the viability of the theory as a whole.

Another thing that I may have failed at is putting sufficient constraints on the kind of triangulation conditions agents can lay out. I mentioned a couple of constraints above. One constraint I imposed, ensures that the g-relation that springs from any set of triangulation conditions will be reflexive. I also imposed a constraint on the extraction of triangulation conditions in section 4.1 involving 'in virtue of' beliefs. But maybe more constraints are required. I do not think, however, that this means the overall theory is not promising. The overall structure of the theory is clear and even if it requires some tweaking, I hope I have spelled out enough of the details to be getting along with.

Despite these limitations, the theory I have presented here is both novel and promising. It may be that there are better, more detailed views that follow the same sort of strategy and I intend to develop the triangulation theory further in later work. But what I hope I have shown is that there is a strategy of which the triangulation theory is an implementation that is promising even if my first attempt to implement the strategy leaves much to be desired.

#### 4.8. Objections and replies

I will now respond to some objections.

##### 4.8.1. The 'too much discretion' objection

At this point, one might object as follows; 'the triangulation theory give agents too much discre-



tion. Surely agents can be wrong about the conditions on subject matter individuation. The triangulation theory is *too* flexible, it ties subject matter to agents' beliefs, which may be mistaken or misguided. The result is that the triangulation theorist cannot distinguish the cases in which an agent is wrong about the conditions on intentional identity from cases in which they are right. There is clearly a sense in which Spinoza, for example, can be wrong about God and, as a result, wrong about the subject matter of his beliefs concerning God. The triangulation theory cannot make sense of this sort of error.'

It is true that nothing in the triangulation theory will yield an answer as to whether or not an agent is right to lay out this or that set of triangulation conditions. But this does *not* mean that, if the triangulation theory is right, the distinction cannot be drawn. We cannot read off the correctness or incorrectness of this or that set of triangulation conditions from the theory of intentional identity. However this *does not* mean that, when it comes to triangulation conditions, anything goes.

There are all sorts of reasons why one might justifiably criticise an agent for laying out this or that set of triangulation conditions. For instance, you might claim that an agent is conceptualising the target wrongly and should adjust their beliefs concerning the target and its place in the world. Or you might mention that other agents understand the putative target and its place in the world differently and, if the agent wishes to co-ordinate and communicate with those other agents, they would do well to line up the triangulation conditions they lay out for the putative target with those of others in her community. In these cases you may be appealing to objective facts about the world. You are claiming that they *should* lay down different triangulation conditions and your claim may be correct. The triangulation theory itself does not yield much guidance on which triangulation conditions an agent *should* lay down, but this does not mean that there is no distinction between correct triangulation conditions that the agent *ought* to lay out and triangulation conditions that they ought not lay out. The standards for correctness and incorrectness come from elsewhere. What the content of an agent's beliefs and what the agent's attitudes are about is one thing, what the content of an agent's belief ought to be and what the agents attitudes ought to be about is quite another.

An example might help illustrate the move I am making. People often debate about torture and what counts as torture. It may be that parties to this debate are aware that their interlocutors are using the word 'torture' differently from themselves. What, then, is the debate about? It is about

how we *should* use the word, what the word *should* mean. This is a serious debate, the stakes are high and the parties to the debate appeal to all sorts of facts to support their claims. It is not as if, just because the parties to the debate are using the relevant notions differently, there is nothing to the debate.

#### 4.8.2. The desire for an objective standard of subject matter individuation

One might object to the triangulation theory as follows; ‘we were interested in the question of when attitudes have a common focus, but you have given us a story about sameness of subject matter relative to a set of conditions. This is a pale substitute for the sort of theory of sameness of subject matter we set out to find, that is, a theory of sameness of subject matter *simpliciter*. For all their faults, at least the views of Salmon, Parsons, Geach and Sainsbury deliver a relation of intentional identity *simpliciter*.’

My response is that the relativity of g-relatedness is independently motivated. The Edelberg-style cases appear to show that either g-relatedness is shifty *or* that the g-relation involved in those cases is non-symmetric. It appears impossible to handle these cases if one has a theory according to which there is just one g-relation that is, in general, symmetric. No theory that delivers this is able to adequately handle Edelberg-style cases. Theories that do involve understanding g-relations as symmetric and non-relative, like those of Geach, Sainsbury, Salmon and Parsons, seem to have great difficulty handling Edelberg-style cases, see section 3.13. The triangulation theorist takes the relativity option, and this relativity is justified by the presence of Edelberg-style cases.

Another reason for the relativisation comes from the fact that, in the hard cases, there appear to be very few worldly resources available to help us decide questions of subject matter. In the hard cases we cannot look to the putative targets themselves (they are missing) or to their place in the world (they do not have one). We are forced to look elsewhere for constraints on intentional identity. If something like the triangulation theory is correct, a good place to look for this guidance is in the details of the beliefs of certain agents. This is something we can do without appealing to worldly guidance on subject matter individuation since all that is required is that agents believe certain things. We were driven to accept a theory that does not appeal to worldly resources because there appear to be very few such resources to which we can appeal in the hard cases. For this reason we should at least be open to considering theories according to which g-relatedness is relativised

and not absolute.

#### 4.8.3. Implicit beliefs and cognitive sophistication

The triangulation theory requires that the relevant agents have fairly sophisticated beliefs concerning the putative target, its place in the world and what it takes for an attitude to be about it. One might worry that this is overly demanding. Often agents do not have this sort of beliefs and it seems that even in the absence of these beliefs we can still correctly ascribe intentional identity.

Here are two forms this sort of objection might take, I will state them in turn. The objection might be put forward in terms of a case in which the relevant agents do not have second order beliefs of this sort and claim that there is, in such a case, intentional identity. Cases involving young children are often wheeled out at this point. Suppose Hob and Nob are four year olds and formed their respective beliefs concerning the witch after overhearing some conversation between their parents. In such a case, says the objector, neither Hob nor Nob plausibly have the beliefs (about what it would take for an attitude to be about particular things) from which we might extract the appropriate triangulation conditions, and yet sentences like (1) may be true.

Another version of this objection runs as follows; even if Hob and Nob are adults it is implausible that they really believe the relevant conditionals that the triangulation theorist says they have to believe if they are to lay down the appropriate conditions.

I have a few things to say in response to this sort of objection. The first involves a clarification of what is meant by ‘belief’ in this context. I am adopting a conception of belief along the lines of that suggested by Lewis (1986, 34-40), Stalnaker (1984, ch. 4,5), and Braddon-Mitchell and Jackson (1996, 187-195). The idea is that an agent’s beliefs are the map by which they steer themselves through the world; beliefs are the things that guide behavioural dispositions. On this conception of belief it seems extremely plausible that adults like Hob and Nob often have the required beliefs. After all, Hob and Nob are disposed to act as if they believe those conditionals and to judge under certain circumstances that this or that attitude is about this or that purported target. They do not, it seems make these judgments arbitrarily. There is plausibly something in their mental states that guides their actions and judgements. A good explanation of this is that Hob and Nob believes the sort of conditionals the triangulation theorist says they do.

But even if one rejects this generous conception of beliefs the triangulation theory is not threatened. All the transparent theory requires is that triangulation conditions are extracted somehow from the psychological state of an agent. If one would prefer to reserve the word ‘belief’ for some other, less common, mental state then the triangulation theorist has a fall back position. The triangulation theorist should say ‘ok, suppose we adopt a more austere conception of beliefs so we can’t read the triangulation conditions that an agent lays out for a putative target off their *beliefs*. The relevant agents are certainly disposed to behave and make judgements in accordance with these conditionals. There is arguably bound to be some mental state or other underlying these behavioural dispositions. We will be able to extract the relevant triangulation conditions from *those* mental states.’

As for the first version of the objection, I admit that the triangulation theorist is committed to there being a point at which an agent becomes sophisticated enough to lay out triangulation conditions and that it may be that this leads the triangulation theory to deliver some somewhat counterintuitive results. This is a bullet I am willing to bite.

I will, however, say two things to suggest that biting this bullet is not too bad. Firstly, as stated above, sometimes the triangulation conditions laid out for the target by the reporter (the person ascribing the attitudes) are invoked rather than those laid out by the agents to which the attitudes are being ascribed. There is, therefore, a reading of (1) according to which it may be true even when Hob and Nob do not lay out any triangulation conditions whatsoever. It may be that when we are assessing cases involving small children there is a tendency to read sentences as invoking the reporter’s conditions. This might explain, and vindicate, the intuition that intentional identity is consistent with the agents to which the attitudes are ascribed being quite unsophisticated.

Another thing to say in response is that, given a generous notion of belief (or the mental state that guides behaviour that an objector might resist calling ‘belief’, see above) from which we are to extract triangulation conditions, it is not implausible that very small children are able to lay out conditions of this sort, though they are likely to be fairly simple. The objector might have had a more austere notion of belief in mind when they presented the children case and once we either adopt a more generous conception of belief or we move to discussing the other sort of mental state which underlies behaviour, it becomes more reasonable to claim that quite young children have

the sort of mental states and resulting pattern of behavioural dispositions from which we can extract triangulation conditions.

#### 4.9. Conclusion

The triangulation theory stacks up well against its rivals. It is more complete than the accounts of Salmon, Parsons, Glick, van Rooy and Geach. It dictates which, if any, causal relations are relevant to intentional identity in any given case and is flexible enough to handle cases involving causal links between the beliefs and cases in which there is no such causal link. The triangulation theorist provides a uniform account of intentional identity, treating cases of intentional identity that involve empty beliefs in the same way as cases involving non-empty beliefs. The triangulation theorist is also in a position to deliver the correct verdicts about Edelberg-style cases, which is more than can be said for the theories of Salmon, Parsons, Geach or Sainsbury. The triangulation theory makes an improvement on Glick's strategy for handling Edelberg-style cases. I have also extended and generalised the theory to handle a wider range of cases than all the alternatives, explicitly accounting for cases involving attitudes with more than once focus, attitudes other than beliefs and so on. For these reasons, the triangulation theory should be considered a strong contender among the extant theories of intentional identity.

## CHAPTER 5: EXPLANATORY BELIEF OR SHARED BELIEF? WE HAVE TO CHOOSE

### 5.1. Two roles for belief

There are at least two roles the notion of belief is supposed to play in our theorising. On the one hand, beliefs are supposed to explain the believer's behavioural dispositions. On the other hand, beliefs are supposed to be shared by different believers and by the same believer across time. That beliefs are shared in this way is often supposed to be central to explaining phenomena such as communication, the stability of belief across time and co-ordination of action. These roles are in tension; by adapting the notion of belief so it can play one of these roles, we compromise its ability to play the other. This chapter has two main goals. The first is to show that there is an interesting tension of this kind and to identify its source. The second is to discuss and evaluate some ways we might attempt to resolve the tension. Once the tension has been made explicit, an intriguing and neglected avenue of inquiry is revealed, in turn suggesting a promising new way of approaching central questions concerning the content of beliefs and other intentional attitudes.

First, I argue that if a theory of belief is going to allow us to correctly explain behavioural dispositions, it must involve understanding singular belief as being sensitive to what information the believer associates with the target of their belief (that is, sensitive to 'how the believer takes the target of their belief to be'). Next, I argue that adopting a theory with this feature means understanding beliefs as extremely fine-grained; small differences in how the believer takes the target of their beliefs to be can make a difference to what is believed. Beliefs, when conceived in this way, appear unsuited to play the *other* role that they are supposed to play in our theorising. Finally, I discuss different moves we might make in the face of this tension, pointing to one promising and hitherto neglected option.

### 5.2. Conflation

One central theoretical role the notion of belief is supposed to play is to help explain the believer's behavioural dispositions. Lewis (1986, 34-40), Stalnaker (1984, ch. 4,5), and Braddon-Mitchell and Jackson (1996, 187-195), among others, adopt this sort of approach to beliefs. The idea is that

agents tend to behave in ways that will bring about what they desire, based on how they take things to be. The ‘how they take things to be’ part of the story is supposed to stand for the agent’s beliefs. An agent’s beliefs are a kind of map that guides their interactions with the world. In this way an agent’s behavioural dispositions are supposed to be closely linked to what they believe.<sup>22</sup>

Consider the case of Lois. Lois works at the Daily Planet. Lois believes that Superman is strong. Unbeknownst to Lois, Superman is identical to Clark Kent, one of her colleagues. Lois wants some help moving furniture and is disposed to ask Superman for help. She sets off in search of Superman, walking past Clark Kent sitting at his desk on the way.

Apparently, Lois believes that Superman is strong but does not believe that Clark Kent is strong. If this is right, we can explain her behavioural dispositions in a straightforward way. Why is she disposed to go off in search of Superman when she wanted help moving furniture? Because she believes that Superman is strong. Why is she disposed to pass by Clark Kent’s desk without stopping to ask for help? Because she does not believe that Clark Kent is strong. If this explanation of her behaviour is going to be available it needs to be one thing to believe that Superman is strong and another thing to believe that Clark Kent is strong.

But according to one theory, often called *the naïve Russellian theory*, to believe that Superman is strong *just is* to believe that Clark Kent is strong. Defenders of naïve Russellianism include McKay (1981), Bealer (1993), Reddam (1982), Salmon (1986, 1989, 2006), Braun (1998, 2001b, a) and Soames (1987) and Tye (1978). Naïve Russellians *conflate* what ought to be considered distinct beliefs, if we are to explain Lois’ behaviour in the natural way. According to the naïve Russellian, singular beliefs are individuated by their content and their content ought to be understood as *Russellian* propositions: pairs consisting of an object, the object the belief is about, and a property, the property ascribed by the belief. If the naïve Russellian theory is right, then the belief that Superman is strong is equivalent to the belief that Clark Kent is strong; they ascribe the same property to the

---

<sup>22</sup> In this chapter I will focus on beliefs. However, since other kinds of intentional attitudes (desires, fears etc.) are also linked to behaviour and are, like beliefs, often supposed to be shared across agents and across time, the tension in question might also be brought out by considering the individuation of these attitudes and their relationship with behavioural dispositions.

same object.

A similar problem arises in cases involving beliefs about collections of objects. Consider the case of Harvey. Harvey is under the impression that some attorneys are not lawyers and he has some friends that he believes to be lawyers and some other friends that he believes to be attorneys. Suppose all and only lawyers are attorneys, but that Harvey is ignorant of this fact. Harvey is broke and wants a loan to make ends meet. He believes that all lawyers are rich. Harvey is disposed to ask his friends who identify as lawyers for a loan. He is not disposed to ask his friends who identify as attorneys for a loan. This case is a variant on a case presented by Mates (1952, 215).

Apparently Harvey believes that all lawyers are rich but does not believe that all attorneys are rich. If this is right, we can explain his behavioural dispositions in a straightforward way. Why is he disposed to ask his friends who he identifies as lawyers for a loan? Because he believes that all lawyers are rich. Why is not disposed to ask his friends who he identifies as attorneys for a loan? Because he does not believe all attorneys are rich. If this explanation of his behaviour is going to be available it needs to be one thing to believe that all lawyers are rich and another thing to believe that all attorneys are rich.

But according to one theory of beliefs, call it the *naïve property theory*, to believe that all lawyers are rich *just is* to believe that all attorneys are rich. There are two forms the naïve property view might take. One involves understanding the content of beliefs in terms of the set of things the belief is about (its extension). Another involves the more general claim that the property of being a lawyer is the same property as the property of being an attorney. I will talk about both of these views under the umbrella of the naïve property view. Whether a naïve property theory is understood extensionally or in terms of the properties involved, it will *conflate* the relevant beliefs, since the belief that Harvey has and the belief that he apparently does not have involve the same pair of properties and the same collection of objects.

The structure of the problem is, in both cases, as follows; it ought to be one thing to believe X and another thing to believe Y but, according to the offending theories, to believe X *just is* to believe Y. Some theories *conflate* beliefs that should, if we are to explain behaviour in the natural way, be understood as different beliefs.

These theories do not make what an agent believes sensitive to the agent's ignorance or error about



the target of their belief. In both cases the agent is ignorant of certain crucial facts about the target of their beliefs. Lois' and Harvey's ignorance is reflected in their behavioural dispositions; Lois acts as if Superman is not Clark Kent and Harvey acts as if it is not the case that all lawyers are attorneys. Since their ignorance *is* reflected in their behavioural dispositions, theories that do not make what agents believe sensitive to their ignorance, like the naïve Russellian and naïve property theories, cannot give the correct explanation of their behavioural dispositions.

The naïve Russellian and naïve property theories deliver a mismatch between what agents believe and their behavioural dispositions. This threatens to undermine our ability to explain the agents' behavioural dispositions by an appeal to what they believe in cases like Lois' and Harvey's. The correct explanation of the agent's behavioural dispositions in these cases requires that we distinguish beliefs that the offending theories conflate.

Note that, as far as the arguments in this chapter go, it does not matter *how* the agent's ignorance is glossed over. Many theories of what determines what an agent believes will deliver the result that the relevant ignorance need not be reflected in the content of the relevant singular beliefs. Some, following Donnellan (1970, 355-356), Putnam (1975), Kripke (1980, 96-97) and others, claim that what singular beliefs an agent has depends on their causal relationship to things in the world and not on any information they associate with the target of their belief. If this is right, agents can pick objects, properties and sets of objects out in their beliefs while being profoundly ignorant about them. According to this kind of causal story, what they believe need not depend on the details of how the believer takes the target of their beliefs to be.

Another theory that allows the relevant ignorance to be glossed over is famously defended by Burge (1986, 1979) and has been defended more recently by Schroeter (2008) and others. According to this theory, facts about the thought and talk in the agent's linguistic community can make a difference to what they believe. For example, what an agent believes is sensitive to what the experts in their community say and think. If this is right, the details of how the agent takes the target of their belief to be need not be appropriately reflected in the singular beliefs they have concerning the target.

One more kind of theory that entails that the relevant ignorance need not be reflected in what the

agent believes are what I call *magnet theories*.<sup>23</sup> According to magnet theories, certain objects and classes of objects are especially eligible to feature in the contents of beliefs. This eligibility is supposed to be a feature of the objects themselves and not dependent on believers or their place in the world. Certain objects and properties are *reference magnets*, especially eligible objects and properties that *attract* beliefs to be about them. These eligibility facts are supposed to make a difference to what an agent believes in a way that need not be reflected in their behaviour or in the information they associate with the target object. If this is right, what singular beliefs an agent, in fact, has need not be reflected in their behavioural dispositions.

### 5.3. The next step

At this point there are two main ways we could jump. The first way is to simply bite the bullet. Claim that, appearances notwithstanding, Lois *does* believe that Clark Kent is strong and Harvey *does* believe that all attorneys are rich. Many Naïve Russellians are tempted by this response to the kind of worries presented above.

Simply accepting this consequence of the naïve Russellian and naïve property theories is an unattractive move. To begin with, there is a widely held intuitive judgment that, in the cases described, Lois does not believe that Clark Kent is strong and Harvey does not believe that all attorneys are rich. Perhaps these judgments can be challenged and our intuitions explained away, but the fact that this result flies in the face of our initial and even considered judgments about the cases gives us reason to be worried about making this move. Secondly, and more importantly, the claim that Lois does believe that Clark Kent is strong is at odds with what appears to be the correct explanation of her behavioural dispositions. Likewise, it is hard to provide a good explanation of Harvey's behavioural dispositions if we are committed to the claim that, in the case described, he believes that all attorneys are rich. If we are to retain the close explanatory link between behavioural dispositions and beliefs, we ought to avoid admitting that the beliefs which appear to be linked to very different behavioural dispositions are, in fact, the same belief.

---

<sup>23</sup> A well-known example of a theory that has this general structure is discussed by Lewis (1983, 370–377, 1984, 227). Similar theories are defended by Sider (2009) and Weatherson (2003)

The other principal option is to adopt a theory of beliefs that *does* make the appropriate distinctions. That is, adopt a theory according to which it is one thing for Lois to believe that Superman is strong and another for Lois to believe that Clark Kent is strong. Similarly, this move involves adopting a theory according to which it is one thing for Harvey to believe that all lawyers are rich and another for him to believe that all attorneys are rich.

Many theories deliver this result. Some well-known examples are *descriptivist theories* of singular beliefs. Descriptivists, such as Russell (1905, 1910-1911) and Lewis (1981, 286-287), tie the *content* of a belief directly to the how the agent takes the target of the belief to be. The descriptivist says Lois' belief that Superman is strong ought to be understood as the belief that the thing that has the properties that Lois associates with Superman is strong. The object the belief is about is not part of the content of the belief. Rather, the idea is that Lois conceptualises Superman in a particular way by associating certain descriptive information with him and it is this associated information that determines the content of her singular belief.

According to the descriptivist theory, the belief that Superman is strong, that Lois has, and the belief that Clark Kent is strong, that she does not have, have different content. One's content is that the thing that has the properties that Lois associates with Superman is strong; the other's is that the thing that has the properties Lois associates with Clark Kent is strong. For the descriptivist, Lois' ignorance *is* reflected in what she believes. The fact that she does not associate the property of being identical with Clark Kent with Superman is reflected in the content of her beliefs.

For the descriptivist, Harvey's beliefs are to be handled in a similar way. Harvey associates certain properties with lawyers and a different set of properties with attorneys. This difference in properties associated with the target explains why the belief that all lawyers are rich, that Harvey has, and the belief that all attorneys are rich, that Harvey does not have, are distinct.

*Fregean theories* like those of Frege (1892), Church (1951) and Chalmers (2002b) also distinguish the relevant beliefs. According to the Fregean, the object, property or set of objects that the belief is about *does* feature in the content of the belief but there is more to the content of beliefs than just the object(s) the belief is about and the property ascribed. There is an extra element, usually some sort of representational entity or mode of presentation, which features in the content of beliefs and allows us to distinguish the relevant beliefs. Different Fregeans say different things about what the

extra element is. Frege (1892) himself calls these entities ‘Sinne’ or ‘senses’. Chalmers and Church refer to the relevant entities as ‘intensions’. But the central Fregean idea is the same: add an element to belief contents so as to make the appropriate distinctions between the beliefs.

For the Fregean, the belief that Superman is strong and the belief that Clark Kent is strong may be distinct even when they ascribe the same property to the same object. Thus, the Fregean delivers the verdict that it is one thing to believe that Superman is strong and another to believe that Clark Kent is strong, as long as these two beliefs involve different senses. The same goes for Harvey’s beliefs.

Another class of theories that distinguish the relevant beliefs in Lois’ case are what we might call *sophisticated Russellian theories*. Descriptivist theories and Fregean theories distinguish the relevant beliefs by distinguishing their *content*. Sophisticated Russellians agree with the naïve Russellian about the *content* of singular beliefs, but still claim that the relevant beliefs are distinct. Let us get clear on the sophisticated Russellian proposal. Consider Perry’s (1979, 5, 18-20) distinction between *what* is believed and *how* it is believed. Two agents may believe the very same proposition in different ways. With this distinction in place the sophisticated Russellian can distinguish the relevant beliefs in Lois’ case; they have the same content but are distinct, because they involve believing that content in different ways.

Sainsbury and Tye (2011) make a move of this kind. They appeal to what they call ‘concepts’. Concepts are to be understood as vehicles of belief and not as forming part of the content of beliefs. According to Sainsbury and Tye, beliefs with the same content may be distinct if they involve different concepts. I will have more to say about Sainsbury and Tye’s account below.

Hopefully this sketch has made the general strategy of the sophisticated Russellian sufficiently clear; the idea is to distinguish the relevant singular beliefs without distinguishing their content. It is hopefully also clear how tools of this kind could be adapted to distinguish beliefs that are about collections of objects. A *sophisticated property theorist*, as we might call them, could claim that the belief that Harvey has and the belief that he appears not to have, have the same content but are nonetheless distinct. They may involve different ways of believing the same content.

All these strategies allow us distinguish beliefs that the naïve Russellian and naïve property theories conflate. Descriptivists and the Fregeans make the distinctions by appealing to differences in

content. Sophisticated Russellians and sophisticated property theorists make the distinctions by appealing to differences in *how* the content is believed. All these theories have implications for the *individuation* of beliefs. We are able to make more distinctions between beliefs and, with these distinctions in place, beliefs are able to better correspond to the believer's behavioural dispositions. The move to individuating beliefs in a more fine-grained way is a step in the right direction, at least as far as correctly linking belief to behavioural dispositions is concerned.

Before we move on, I want to note that the issue is *not* primarily one of the semantics or pragmatics of belief reports. We want to be able to explain the relevant behavioural dispositions in the cases described above. This guides our choices between theories of belief quite independently of our preferred theory of the semantics or pragmatics of belief reports in natural languages. No amount of fiddling around with the pragmatics or semantics of belief reports will yield a solution to the issue raised above; it is an issue concerning beliefs themselves not the sentences we use to report them. It may be that, when faced with some *other* challenge,<sup>24</sup> we will be forced to adopt a complicated semantics or pragmatics of belief reports but this move would do little to answer the challenge at hand.

#### 5.4. A cautionary tale

If we are going to make the appropriate distinctions between beliefs so as to line beliefs up with behavioural dispositions in the right way, we must be careful about *how* we spell out the details of the proposals discussed in the last section. One way of identifying and avoiding pitfalls is to consider one account that that goes wrong and to learn from its mistakes. In this section I will provide a more detailed sketch of the account defended by Sainsbury and Tye, and discuss how and why it goes wrong. I will then extract a moral about how we *should* implement the above strategies, if we are concerned with correctly explaining the relevant agent's behavioural dispositions in the cases described.

Sainsbury and Tye (2011, 101) divide concepts into two kinds, 'nominative concepts' that refer to particular objects, and 'predicative concepts' that refer to properties and, indirectly, to the collection of objects that have those properties. Some concepts do not pick out any object or class of

---

<sup>24</sup> For a clear discussion of some such challenges see Saul (1997, 1999).

objects.

Sainsbury and Tye (2011, 102-105) also divide *uses* of concepts into two sorts, originating uses and non-originating uses. Non-originating uses involve deference to other uses of the same subject in the past or uses of other subjects. Originating uses are the ones that *do not* involve this kind of deference. Originating uses bring concepts into existence. This can be done in a number of ways but the most obvious way is when someone explicitly introduces a concept, e.g. as part of a novel scientific theory or so as to pick out a particular object, like a person.

The idea is that there are chains of non-originating uses of a concept leading back its originating use. Non-originating uses of a concept involve the concept they do in virtue of these causal chains of deference leading back to the concept's originating use. For every concept there is just one originating use (Sainsbury and Tye 2011, 104). Concepts are wholly individuated by their originating use; concepts are identical if, and only if, they have the same originating use (Sainsbury and Tye 2011, 105-106).

Sainsbury and Tye briefly discuss the relationship between the content of a concept and the concept itself. The 'content' of a concept, for Sainsbury and Tye, is identified with its referent. The argument against the Sainsbury and Tye view given below will not turn on the details of the relationship between originating use of a concept and its referent.

Now consider a case, due to Kripke (1979, 449), to which I have added some details. Suppose a child is born and his mother names him 'Paderewski', introducing the concept PADEREWSKI. With all her subsequent uses of the concept she intends to use the concept as she used it before. Paderewski grows up and becomes both an accomplished pianist and a politician. One night, Paderewski is playing piano at a jazz concert and an audience member, Peter, is impressed by the performance. Peter collects signatures of accomplished pianists and he believes, in light of the excellent performance, that the pianist, introduced as 'Paderewski', is an accomplished pianist. He defers to others' use of the concept PADEREWSKI such that, by Sainsbury and Tye's lights, his belief involves the concept that was introduced by Paderewski's mother. Peter is disposed to ask Paderewski for his autograph but he does not get the opportunity as Paderewski leaves the venue mere minutes after the end of the concert.

The next day Peter notices a political rally taking place near his home. He listens to a speech

delivered by a politician who is introduced as 'Paderewski'. Peter does not recognise that the politician and the pianist from the night before are the same person. Peter believes that no politician is an accomplished pianist, so he comes to believe that Paderewski, the politician, is *not* an accomplished pianist but that he is an excellent speaker. He defers to others' use of the concept PADEREWSKI such that, by Sainsbury and Tye's lights, his belief involves the concept that was introduced by Paderewski's mother. Peter is *not* disposed to hang around the political rally, hoping to get an autograph. He goes home, which is a shame; there were ample opportunities for him to get Paderewski's autograph at the rally.

Does Peter believe that Paderewski, the man who gave the speech, is an accomplished pianist? It appears as if the correct explanation of his behavioural dispositions requires that he does not. He is disposed to hang around the venue after the concert in an attempt to get Paderewski's autograph. This is nicely explained by his believing that Paderewski, the pianist, is an accomplished pianist. He is *not* disposed to hang around the rally and attempt to get his autograph, even though he desires the autograph of an accomplished pianist. The natural explanation is that he does not believe that Paderewski, the politician, is an accomplished pianist. If he did, he would presumably be disposed to hang around the political rally in the hopes of getting his autograph.

But according to the Sainsbury and Tye view, these beliefs are equivalent. Not only do they ascribe the same property to the same object, they also involve the very same concept. The concept involved in these two beliefs has the same originating use, so, by Sainsbury and Tye's lights, the beliefs involve the very same concept. The Sainsbury and Tye theory still conflates beliefs that ought to be considered distinct.

What went wrong? I submit that the Sainsbury and Tye theory goes wrong for the same reason the naïve Russellian and naïve property theories go wrong. The Sainsbury and Tye theory does not require that the agent's ignorance or error regarding the target of their belief and its place in the causal order of things be reflected, in the appropriate way, in which singular beliefs they have.

Peter is ignorant about the object his belief is about *and* he is in error about the concept he employs. This ignorance and error is reflected in his behaviour. Sainsbury and Tye's theory glosses over this ignorance. The problem is that these two beliefs can involve the same concept without this fact being reflected in Harvey's behavioural dispositions. Sainsbury and Tye do not link beliefs to

behaviour in the appropriate way. They still do not make *enough* distinctions between beliefs.

Whichever of the strategies sketched in the last section we adopt, we ought to individuate the relevant singular beliefs in such a way that their singular belief is sensitive to what descriptive information the agent associates with the target, which would encode things like their ignorance and error regarding the target. Cases like Peter's show that Sainsbury and Tye do not carve beliefs finely enough, if our goal is to explain Lois' behaviour in terms of her singular beliefs.

This same problem will arise for some Fregean theories. Frege and Fregeans are not often explicit about how senses come to be involved in the content of beliefs; they are not explicit about what it takes for an agent to *grasp* a sense.<sup>25</sup> But whatever story the Fregean tells about how beliefs come to involve the sense they do, if two beliefs can come to involve the same sense without that fact being reflected in the agent's behaviour, then it looks as though cases like Peter's will be possible. We will be able to cook up scenarios at which two beliefs, in fact, involve the same sense and reference but at which the believer will not act accordingly.

How can we avoid this pitfall? The answer, I think, is to link what is believed more closely to the descriptive information that agents associate with the target of their belief. Descriptivists, who tie the content of singular beliefs directly to how the agent takes the target to be, have nothing to fear from cases like Peter's. Neither do Fregeans who link which sense is involved in the content of a belief directly to how the believer takes the target to be, nor do sophisticated Russellians or property theorists who link concepts or ways of believing propositions to how the believer takes the target to be. It is when we adopt an account of belief that glosses over the details of how the believer takes the target of their belief to be that we run into trouble. The moral of the cautionary tale is this: if we are going to avoid giving rise to conflation altogether, we had better understand what an agent believes as being appropriately sensitive to the details of how they take the target of their belief to be.

---

<sup>25</sup> Although Frege is not explicit about this issue he does say things that constrain what story he can tell about grasping senses, if he wants to stay consistent. For instance, he claims that two agents can grasp a sense while disagreeing about which descriptions apply to the target entity. See Frege (1892, n. 2).



### 5.5. The cost of explaining behaviour

The other role beliefs are supposed to play seems to require that beliefs are shared, in some crucial sense, across agents and across time in a way that ignorance is not. When two agents believe that Jill is a philosopher, there is supposed to be some crucial sense in which they have the same belief. Likewise, in intra-personal cases, there is supposed to be a crucial sense in which an agent who believes that Jill is a philosopher on Monday and believes that Jill is a philosopher on Wednesday has the very same belief on Monday and Wednesday. Beliefs appear to be *shared* across agents and across times.

Another reason one might want beliefs to be shared concerns a plausible theory of successful communication. Stalnaker (1978, 1988, 151) defends an elegant and, *prima facie*, plausible account of successful communication which we might call, borrowing some terminology from Weber (2013, 207-208), *the FedEx model of communication*. The FedEx model says that successful communication involves an agent expressing a belief and their audience coming to have that very same belief. Suppose you believe that Jill is ill and want to tell me about it, so you say ‘Jill is ill’. According to the FedEx model, the communicative act is successful if I come to believe the same thing that you believe, that Jill is ill. The belief that the audience comes to believe is supposed to be identical to the belief expressed by the speaker, at least in the good case.

But if the content of an agent’s belief regarding the target is sensitive to the details of how they take the target to be, the claim that beliefs are shared is threatened. What information is associated with the target is *not* shared in this way. It will very seldom be that two agents conceptualise the target of their beliefs in *exactly* the same way. If we tie what singular beliefs an agent has to how the agent takes the target to be, singular beliefs become problematically difficult to share. In this way a tension arises, let me illustrate.

Suppose two agents believe that Jill is a philosopher. These agents likely have slightly different beliefs about Jill. Perhaps one of the agents knows Jill a little better and has a few more beliefs about Jill. If our theory of beliefs tells us that any differences in how an agent takes the target of their singular belief to be can make a difference to what singular beliefs an agent has about that target, then we are pushed to admit that, appearances notwithstanding, these agents may not, in fact, have the same belief. What they believe is tied to how they characterise the target *and* they

characterise the target differently.

Next consider an intra-personal case. Suppose on Monday an agent believes that Jill is a Philosopher, on Tuesday she comes to believe that Jill was born in Chicago, on Wednesday she still believes that Jill is a Philosopher. Is the belief that the agent has on Monday the same belief as the one she has on Wednesday? The way she takes Jill to be on Monday is different from how she takes Jill to be on Wednesday. If our theory of beliefs tell us that these differences can make a difference to what is believed then we are pushed to admit that the Monday belief and the Wednesday belief may not be the same belief.

A similar concern arises regarding communication. If beliefs are not often shared, then successful communication, as understood in terms of something like the FedEx model of communication, will be extremely uncommon. The general worry is that if agents don't share beliefs then certain communicative phenomena appear difficult to explain.

Whatever the faults of the naïve Russellian and naïve property theories, at least they allow the relevant beliefs to be often shared by agents and by an agent across time. By glossing over certain facts about how the believer takes the target to be, they gloss over differences between agents that might, according to other accounts, undermine the claim that the beliefs in question are shared between agents and across time.

Allowing beliefs to be shared is also an appealing feature of certain Fregean theories, sophisticated Russellian and sophisticated property views. The idea is to carve beliefs coarsely enough such that agents often share beliefs, while still making enough distinctions to distinguish the beliefs in cases like Lois' and Harvey's. Sainsbury and Tye (2012, 21-22, 24) claim that concepts are shared across agents. This is one of the main reasons they mention *not* to tie concept individuation to the agent's beliefs about the world. Instead Sainsbury and Tye, as we have seen, tie belief individuation to certain causal facts. Some idiosyncrasies of how the agent takes things to be are glossed over and this allows for shared belief.

A similar thought appears to have motivated Frege's claim that two agents can employ the same sense while associating different descriptions with the target. Frege (1892, 26-27, n.2) and some Fregeans want to resist tying which sense is involved in a belief directly to what properties the believer associates with the target. Frege (1892, 25-27) and some Fregeans think, rightly if you

ask me, that the claim that senses are individuated by an associated description is hard to reconcile with the claim that beliefs are often transmitted from one agent to another. But if the Fregean severs the link between how the agent takes the target of their beliefs to be and what they believe, then they will have trouble correctly explaining behaviour. We will be able to cook up cases like the Paderewski case described in section 5.4 that spell trouble for Fregeans who sever this link. This sort of Fregean view is a good example of a theory caught in the middle of the tension I have in mind. On the one hand, they want senses to be linked to the inferential import and cognitive significance of beliefs and since being disposed to make certain inferences is a kind of behavioural disposition, sense individuation is also supposed to be linked to behavioural dispositions. On the other hand, they want senses to be individuated in such a way that they are shared across agents and across time; senses are supposed to be public, in some sense. They cannot have it both ways, it seems.

There is a tension between two roles the notion of belief is supposed to play. As we make more and more fine-grained distinctions between beliefs so as to be able to correctly link what an agent believes to their behavioural dispositions, we make beliefs more and more seldom shared. We then have trouble explaining phenomena that appear to require beliefs to be often shared by different agents and by the same agent across time.

## 5.6. Options

There are three main moves we might make when faced with this tension. The first, discussed in section 5.4, is what I call the settling option. The second involves adopting a pluralist theory of beliefs. The third involves learning to live with extremely fine-grained beliefs by giving up on the orthodox understanding of shared belief and related phenomena.

### 5.6.1. Settling

The settling option involves sacrificing some of our ability to explain behavioural dispositions in terms of the agent's beliefs at the altar of shared belief. The claim that beliefs are shared between agents and across time is plausible and attractive. Attractive enough, perhaps, to motivate us to compromise the explanatory link between behavioural dispositions and belief.

Once we see the tension it becomes clear why some find it appealing to simply bite the bullet when faced with cases like Lois' and Harvey's. The thought, expressed by Kripke (1979, n.15, n.23, n.26 and n.28) and Salmon (1989, 1990), is that if the argumentative strategy outlined in sections 5.2 through 5.4 inclusive is applied in its full generality, beliefs will become *unacceptably* fine-grained.<sup>26</sup> We might be tempted to nip the argument in the bud and stop the fine-graining strategy at this low level of granularity.

In so far as we want to retain shared belief by compromising our ability to explain certain behaviour we will commit ourselves to biting some bullets in cases like Lois', Harvey's or Peter's. The different settlers may disagree on which bullets to bite but they are all going to have to bite some bullets somewhere or other. Compromising the link between behavioural dispositions and beliefs in these cases is a serious cost but one that some may be willing to accept.

This option is costly for another related reason; if what agents believe in cases like those discussed above does not explain the difference in behaviour, something else will have to. The fine-grained behavioural dispositions need to be explained somehow. If we give up on explaining the relevant behavioural dispositions by appealing to what the agent believes, we will need to develop some *other* explanatory resources to do the work.

### 5.6.2. Pluralism

A second way we might try to resolve the tension is by adopting a pluralist theory of belief. Perhaps there are two notions of belief or two kinds of belief content in play, one of which is closely linked to explaining behavioural dispositions and another which is shared across agents and linked to related phenomena. Chalmers (2002a) puts forward one theory of this sort.

I will make two remarks about the move toward pluralism. The first is a complaint about Chalmers's particular pluralist proposal. The second is a more general word of caution about attempting to resolve the tension by adopting a pluralist theory of belief.

---

<sup>26</sup> The concern is more often put in terms of synonymy of linguistic terms but this can be understood as an instance of a more general worry about stability.

Chalmers suggests that beliefs have two kinds of content: epistemic content and subjunctive content. Epistemic content is tied to the believer's behavioural dispositions, and the epistemic content of a belief is supposed to capture how the believer take things to be, relative to their place in the world. According to Chalmers (2002a, 608-609, 611-612), the epistemic content of a belief is wholly determined by the internal state of the believer. The epistemic content of an agent's belief captures how they take things to be, relative to their (the believer's) place in the world. The epistemic content of beliefs is captured by a set of *centered worlds* (Chalmers 2002a, 611). A centered world is a possible world with a mark that indicates an individual at a time that serves as the center of that world. More formally, a centred world is an ordered triple composed of a world, an individual, and a time. The center of these worlds represents the believer's perspective. To take an example, the epistemic content of my belief that Aristotle was a great philosopher would be captured by a set of centered worlds in which there is an object that has the properties I associate with Aristotle (which includes the property of standing in a particular relation to me) and was a great philosopher.

Subjunctive content is much less fine-grained. The subjunctive content of a belief, depends on the internal state of the agent *and* on how the world turns out to be. The idea is that the epistemic intension of a belief determines its extension at the actual world and the subjunctive content of that belief concerns *that object*. For example, the subjunctive content of my belief that Aristotle was a great philosopher can be captured by the set of worlds at which the object that has the properties I associate with Aristotle and stands in a particular relation to me in the actual world, was a great philosopher. Understood in this way, subjunctive content will be often shared, and shared subjunctive content is supposed to help explain communication and co-ordination of belief and action (Chalmers 2002a, 631 n.29). By linking one notion of belief to epistemic content and the other to subjunctive content, Chalmers is thus able to resolve the tension between the two roles belief is supposed to play.

The complaint against Chalmers's proposal is that his subjunctive content does not explain what he wants it to explain. According to Chalmers's story, communication and co-ordination of action often involve shared subjunctive content between the audience and the speaker. When a speaker says something like 'Jill is ill' the audience acquires a belief whose epistemic content is different from the speaker's, but whose subjunctive content is the same, at least when the communicative

act is successful.

The problem is that successful communication often occurs even when, by Chalmers's lights, the relevant beliefs have no non-trivial subjunctive content. Chalmers discusses a case in which Karen is agitated on Christmas Eve because she believes that Santa is coming. The epistemic content of her belief explains her behaviour. But, says Chalmers (2002a, 623), her beliefs have no non-trivial subjunctive content; they fail to pick out any object in the actual world, as there is no Santa Claus.

Consider a case in which two detectives investigate the death of Smith. They come to believe that Smith was murdered and begin forming beliefs about Smith's murderer, whom they dub 'mad Mel'. As it happens, Smith's death was an accident, there is no such person as mad Mel. So the detectives' respective beliefs about mad Mel have no non-trivial subjunctive content.

Yet the detectives may share beliefs about mad Mel. They might share the belief that mad Mel is tall. They may also communicate about mad Mel, sharing their thoughts about how the murder was committed. They can co-ordinate their action toward finding and catching mad Mel. They can agree and disagree about her. It might be that one detective thinks she is still in town while the other thinks that she skipped town the first chance she got. Whatever this sort of communication, disagreement and shared belief involves, it does not, by Chalmers's lights, involve shared subjunctive content since the beliefs in question have no non-trivial subjunctive content.

What is more, some beliefs are shared across time in a way that does not depend on shared subjunctive content. Suppose, on Monday Charlie, one of the detectives, believes that mad Mel is evil. On Tuesday Charlie comes to believe that mad Mel is also cunning. On Wednesday Charlie may still believe what she believed on Monday, that mad Mel is evil. In this case, these beliefs do not have any non-trivial subjunctive content; mad Mel doesn't exist. Charlie's Monday belief and her Wednesday belief appear to be the very same belief, in some crucial sense. But in what sense are they the same belief on Chalmers's picture? They differ in their epistemic content and neither belief has any non-trivial subjunctive content.

Why is this bad news for a defender of Chalmers's proposal? Because subjunctive content may, before long, be out of a job. Assuming that the communication and shared belief in the detective case is *explicable*, there will be a correct theory of communication and of the sense in which beliefs are shared in such cases. This theory will explain the communication and shared beliefs without

appealing to shared subjunctive contents. But once this theory is in place, why should we only apply it to cases in which the beliefs have no subjunctive content? What is stopping us explaining *all* communication and shared belief in that way? Nothing, as far as I can see. If we can explain all communication and shared belief in this way, one main reason to adopt a pluralist theory and to posit a second dimension of content is undermined. The detective case also teaches us something about *how* the pluralist option ought to be pursued. It had better be that the kind of belief or belief content that is supposed to be shared across agents and across time and is supposed to explain communication, is present in cases in which the relevant beliefs have no extension at the believer's world.

If epistemic content were supposed to be derived from, or in some way secondary to, subjunctive content, it would then be easy to explain why we need to appeal to subjunctive content, even if the epistemic content does most of the work explaining these phenomena. But Chalmers (2002a, 621-622) explicitly argues that epistemic content is primary and that subjunctive content is derivative. Given this part of Chalmers's view, he cannot consistently denigrate epistemic content in the required way.

The second point I want to make concerning pluralism is more general, and applies both to Chalmers's pluralism and to other possible implementations of a pluralist approach. When confronted by the tension discussed above, it is tempting to claim that there are, in fact, two kinds of belief in play or that the term 'belief' is ambiguous. We can then claim that what was apparently a tension was just evidence of a pluralism, either in kinds of beliefs or in readings of the word 'belief'. But the very ease of this move ought to make us cautious of adopting it. When faced with counterexamples, theoretical tensions or counterintuitive results, a defender of a theory can almost *always* appeal to the claim that there are, in fact, two sorts of things in play or to posit an ambiguity in one of the crucial theoretical notions. If this move is available to all participants in a discussion we ought to worry that it is not as respectable a panacea as it might appear to be. We should *only* posit an extra theoretical kind or appeal to an ambiguity in a theoretical term if we really need to. Here I partially echo some remarks of Kripke (1977, 268). This is a general methodological reason to prefer, all else being equal, a non-pluralist theory according to which there is just one notion of belief and one kind of belief content in play. It may be that all else is not equal; it may be that once we discuss the other options for resolving this tension in enough detail they will turn out to be

limited in crucial ways. We may then be forced to adopt a pluralist theory. However, until we have carefully evaluated the other options, we should at least be cautious about helping ourselves to pluralist resources in the interests of resolving the tension outlined above.

### 5.6.3. Off the well beaten track

Now for the third option that I favour and which, as far as I know, has been neglected in the literature. The stability of belief and related phenomena have often been understood in a particular way: in terms of *shared* belief, i.e. in terms of the agents in question having a belief of exactly the same type. In light of the tension, we might resist this conception and seek an alternative account of stability and related phenomena. After all, beliefs stand in all sorts of relations other than the identity relation. Is it so implausible that some of these other relations might explain the relevant phenomena?

Those who wish to make this move might claim that ‘believing the same thing’ in the relevant sense need not involve sharing a belief, rather it involves having beliefs that stand in some other kind of relation. For instance, we might say that Jack believes the same thing as Jill when Jack’s belief is related to Jill’s in a certain way. Similarly, communication could be explained in terms of the parties to the communicative interaction having suitably related beliefs.<sup>27</sup>

Naturally anyone who adopts this strategy will have to provide a detailed account of just what this other relation consists in, which is no easy task. But this strategy, if it could be made to work, would have some major advantages over its rivals. We would be able to carve beliefs finely enough so as to allow for the correct explanation of behavioural dispositions in troubling cases, while at the same time retaining our ability to explain phenomena like communication and stability of belief across agents and across time.<sup>28</sup> By understanding the phenomena of communication and stability

---

<sup>27</sup> Another interesting question is whether a similar move can be made in explaining the stability of other intentional attitudes across agents and across time. For example, it may be that my desire for food and your desire for food are not *the same desire* but are, nonetheless suitably related so that the relevant sameness of desire claims come out true.

<sup>28</sup> Weber (2013) sketches a theory according to which shared belief is not a necessary condition for successful communication. Weber is motivated by considerations to do with centred belief and



of belief in a new way we might be able to live with extremely fine-grained and idiosyncratic beliefs. I will discuss one way adopting this sort of strategy in the next chapter.

### 5.7. Conclusion

The arguments presented above suggest that, so long as we hold onto the understanding of stability that involves *identity* of belief, some sort of trade-off is inevitable. We must either compromise the explanatory link between beliefs and behavioural dispositions, claim that beliefs are not often shared, admit that there is more than one notion of belief in play or reject the conception of stability of belief that relies on beliefs being shared. In this chapter I have made the tension explicit, provided a diagnosis of why it arises and suggested a few options available to us when faced with the tension. Understanding the tension in the way described above indicates a potentially fruitful avenue of inquiry.

There are good reasons to avoid the settling option. It flies in the face of intuitive judgments about what agents believe, it means loosening the explanatory connection between beliefs and behavioural dispositions, and it means that we would have to find some *other* resources to explain the relevant behavioural dispositions.

The pluralist approach is *more* promising than the settling approach. If we adopt a pluralist approach we can appeal to the relevant beliefs to explain the relevant behaviour while at the same time admitting that there is a kind of belief or belief content that is shared across agents. I have provided reasons to reject Chalmers's particular pluralist proposal and have presented some general reasons to be cautious of adopting the pluralist approach.

The most promising way forward is, I think, the third option. If stability of belief is understood in terms of some relation between beliefs other than strict identity, we can neatly avoid the costs of the settling approach while, at the same time, accounting for stability and related phenomena. This option is not only promising but is also, for the most part, unexplored.

---

not by the sorts of cases above. His theory is an example of how theories of communication that do not require identity of belief might go.



## CHAPTER 6: PUTTING THE TRIANGULATION THEORY TO WORK

### 6.1. An unorthodox approach

In this chapter I will defend a package of views composed of a descriptivist theory of the content of attitudes and a triangulation theoretic account of subject matter. In addition to the independent merit of the triangulation theory as a theory of subject matter, triangulation theoretic tools allow descriptivists, such as Lewis (1981) and Jackson (2010), to respond to a whole swath of objections. With the help of the triangulation theory of subject matter, we can keep many of the benefits of a descriptivist theory of content while, at the same time, avoiding many of its apparent costs. My strategy centres on driving a wedge between questions of subject matter and questions of content.

This chapter will be split into three parts. In the first part, sections 6.2 and 6.3, I outline the descriptivist theory, spell out how descriptivism and the triangulation theory fit together, and discuss one of the most serious objections levelled against descriptivist theories of attitude content. What I have to say in response to this objection will bring out a central feature of the package of views I defend.

In the second part, sections 6.5 through 6.9 inclusive, I implement the strategy discussed in section 5.6.3. By developing a new understanding of certain phenomena traditionally associated with agents' having the same attitude, we are able to hold onto the claim that attitudes are extremely fine-grained while, at the same time, giving an acceptable account of these phenomena.

In the third part, sections 6.11.1 through 6.11.4, I respond to some other arguments against the descriptivist theory of the content of attitudes. There are objections to the description theory of names that many believe undermine both that theory *and* its cousin, the descriptivist theory of attitude contents. I will discuss how one might transform some of the traditional arguments against the description theory of names into arguments that target descriptivism about the content of attitudes. I will argue that the arguments, thus transformed, fail to undermine the descriptivist theory of attitude content.

### 6.2. The descriptivist theory of the content of attitudes

I will be defending a descriptivist theory of attitude contents of the kind defended by Quine (1948),

Lewis (1981, 286-287), and others. The idea behind the descriptivist theory of attitude contents is that attitudes should be understood in terms of the some of the properties that agents associate with the putative target of their attitudes. For example, the content of attitudes like the belief that London is pretty should be understood as the belief that there is a thing that has a certain set of properties (the ones the believer associates with London) and that that thing is pretty. More formally, the content of the belief is that ‘for some object x, x has properties F, G, H...<sup>29</sup> and x is pretty’. This belief gets to be true just in case there *is* an object that has those properties and is pretty. The object that happens to have these properties does not feature in the content of the belief. A similar story will be told about fears, desires, etc.

The descriptivist theory of attitude content has some very attractive features. As I discussed in the previous chapter, descriptivism allows us to make crucial distinctions between attitudes that some other theories do not make. In the case of Lois, if we want to explain Lois’ behaviour in the natural and intuitive way there is pressure to say that Lois believes that Superman can fly but *does not* believe that Clark Kent can fly. The descriptivist delivers the result that it is one thing for Lois to believe that Superman can fly and another thing for her to believe that Clark Kent can fly. The descriptivist will say that when Lois believes that Superman can fly, the content of her belief is that there is a thing that has the properties that Lois associates with Superman and that thing can fly. Whereas the content of the belief that she would have if she believed that Clark Kent can fly would be that there is a thing that has the properties Lois associates with Clark Kent and that thing can fly. The beliefs involve different associated properties and therefore, by the descriptivist’s lights, have different content.

There is pressure to think that it is one thing to believe that Hesperus is Phosphorous and another to believe that Hesperus is Hesperus. The descriptivist delivers this distinction. The content of the first belief is that the thing that has the properties the believer associates with Hesperus is the thing that has the properties the believer associates with Phosphorous. The content of the second is that the thing that has the properties the believer associates with Hesperus is the thing that has the

---

<sup>29</sup> The set of properties an agent associates with the putative target sometimes includes uniqueness properties. An agent might think that the putative target of their attitude has certain properties *and* has the property of being the only thing with those properties.

properties the believer associates with Hesperus.

Descriptivism also yields an attractive account of the content of empty attitudes. Suppose I believe that Pegasus does not exist. It appears as if this belief has content and is true. But if we think that objects feature in the content of beliefs of this sort, we are in a bind. In this case, we are faced with two options. Either we are forced to claim that the attitude does not have content, since there is no object *Pegasus* that can feature in the content of the attitude, in which case it becomes hard to explain how the belief seems to be both meaningful and non-trivially true. Or we must commit ourselves to an object, Pegasus, of which the belief says that it does not exist,<sup>30</sup> committing ourselves to some sort of mythical, non-existent or merely possible object. Either way does not look terribly attractive. These options are two of the most obvious approaches to the problem of true negative existential beliefs, but they are by no means the only options available. If a descriptivist theory is correct, objects like Pegasus does not feature in the content of beliefs of this sort. The content of this empty belief is that there is nothing that has a certain set of properties, the properties that I associate with Pegasus, the putative target of my belief. On this understanding my belief has content and is, presumably, true.

Descriptivism delivers a uniform treatment of the content of empty and non-empty attitudes. In both kinds of cases it is the associated properties, and not the target object itself, that feature in the content of attitudes. So the mere fact that the object appears to be missing when the attitudes are empty makes no difference to how the descriptivist handles the content of attitudes. The promise of a uniform treatment of content is an attractive feature of descriptivism. Despite these attractions, the descriptivist theory of attitude contents faces some serious and familiar challenges.

The first complaint against the descriptivist that I will discuss says that the descriptivist misclassifies beliefs about particular objects as general quantificational beliefs. Call this *the generality com-*

---

<sup>30</sup> Strictly speaking those who take this route and commit themselves to mythical or merely possible objects cannot say that these things don't exist, since they either exist at some possible world or exist as abstracta at the actual world. Those who take this path must also interpret the belief, not as ascribing non-existence to an object, but rather as ascribing non-concreteness or non-actuality to an object. The need to reinterpret beliefs in this way is another cost of these approaches.

*plaint*. The complaint goes something like this: ‘according to the descriptivist, the belief that London is pretty should be understood as a general belief about what there is: the belief that, for some  $x$ ,  $x$  has the London properties and  $x$  is pretty. But this is a mistake, the belief’s being about a particular thing has been washed out by the descriptivist analysis. We started with an attitude about a particular object and ended up with a general belief about what there is. General beliefs are not directed at a particular object. So something is wrong with the descriptivist conception of attitude content.’

The generality complaint and my response to it will bring out an important feature of the package of views I am defending. I will, therefore, discuss this objection and my response to it in some detail before moving on to answering some other objections levelled against the descriptivist theory of attitude content.

### 6.3. Content and subject matter

Once we combine a descriptivist theory with a triangulation theory of subject matter, we can answer the generality complaint. With the introduction of triangulation theoretic tools, the descriptivist is able to drive a wedge between questions of content and questions of subject matter. Descriptivism is a theory of *content* and the triangulation theory yields an account of *subject matter*.

The idea that we can read the subject matter of an attitude directly off its content is implicit in much of the debate about intentional attitudes. Consider, for instance, a naïve Russellian account of content according to which the object that the attitude is about forms part of the content of that attitude. Suppose that Zaphod and Trillian both believe that Ford is happy. For the naïve Russellian, the content of Zaphod’s belief that Ford is happy is to be captured by a proposition composed of Ford, the man himself, and the property happiness. For the naïve Russellian, the object the belief is *about* is right there in the content. So when we want to know what an attitude is about we need simply look at which objects feature in that attitude’s content. What is more, if we know the Russellian content of two attitudes, we can determine whether or not they have the same subject matter without any further information. Trillian’s belief and Zaphod’s belief have the same subject matter because Ford, the man himself, shows up in both of the attitude’s Russellian content. An extremely natural and prima facie attractive package of views consists of naïve Russellianism about content and a transparent theory of intentional identity. Salmon (1989, 369, 2005, 105–108), for instance,

is both a naïve Russellian about the content of attitudes and a defender of a transparent theory of intentional identity. If we adopt this package of views, we can read off sameness of subject matter from the content of pairs of beliefs.

According to this way of thinking about the relationship between subject matter and content, if we know *what* is believed, feared, desired, etc., we will be in a position to know which object or objects these attitudes are about. Understanding subject matter in this way also yields a straightforward test for disagreement; we can say that two beliefs involve disagreement about the target if one object, concept or aspect features in the content of both beliefs and they ascribe conflicting properties. If we think that the content of a belief fully determines its subject matter, we can link questions of subject matter, content, agreement, and disagreement together closely.

This way of understanding the relationship between content and subject matter is presumably behind the generality complaint. If the *content* of a belief is general and we can read off the subject matter of an attitude from its content, it must have general, rather than particular, *subject matter*.

The complaint seems to hinge on the claim that the subject matter of an attitude is, in general, reflected in its content. I suggest that the descriptivist should respond by denying this claim. If she adopts a triangulation theoretic approach to subject matter and combines it with a descriptivist theory of content she can say that, even though the attitudes have general, quantified content they still get to be *about* particular putative targets. Triangulation theoretic tools allow the descriptivist to tell a story about the subject matter of attitudes (their putative target, what they are ‘about’) that is consistent with construing attitude content as general. The belief that London is pretty might be about London, the putative target of the attitude, *and* have the kind of general content the descriptivist claims it does. London, the putative target as conceived by the relevant agent, is what the belief is about. The putative target need not feature in the content of the attitude.

To take another example, suppose that there is a large animal in the lower paddock and Jimmy is afraid of it. In what sense is Jimmy’s fear about the large animal? The triangulation theorist ought to say that Jimmy’s fear is about the animal because Jimmy has a conception of the putative target of his fear and he has beliefs about what it would take for attitudes to be about that putative target. From these beliefs we can extract triangulation conditions. If Jimmy’s fear satisfies at least one of these conditions then, by Jimmy’s standards, his fear focuses on the large animal in the paddock.

An objector might say, at this point, that I have dodged the question. The question was: in what sense is Jimmy's fear about the large animal? The resources supplied might help settle whether or not Jimmy's fear is about the thing that Jimmy takes to be in the lower paddock, but that does not seem to bear on the original question.

This kind of objection brings out an important point. The triangulation theoretic tools I have developed do not pertain to what we might call the *reference* of a belief (as opposed to its subject matter and its content). One might think that the *reference* relation obtains only when there really is an object that the relevant attitude picks out. Empty attitudes can have subject matter in spite of, apparently having no reference. I will put aside discussion of the reference relation for the purposes of this chapter and this thesis. I am concerning myself with questions of subject matter and content. In future work I plan to discuss the relationship between *reference*, content and subject matter.

There is one, slightly cheeky, thing that I can say in response to this kind of objector. I might say the following: 'the triangulation theory can determine whether or not Jimmy's attitude is directed at the thing in the paddock that that *you*, the objector, are thinking about when you consider the case. So there is a sense in which I can answer the question "in what sense is Jimmy's fear about the large animal?"' The answer is that, by the relevant triangulation conditions, Jimmy's attitude has the thing that the presenter of the case had in mind when they set out the case as its subject matter. Maybe the objector will complain that I have not taken the question in the spirit in which it was asked. The objector might have had in mind the relation of *reference* discussed in the previous paragraph. Certainly the triangulation theory does not deliver an account of *that* relation, but part of what I am trying to bring out in this thesis is that the relations of *reference*, in this sense, do not appear to play any distinctive role in determining the subject matter of intentional attitudes.

As we will see below, in allowing the descriptivist to handle content and subject matter separately, triangulation theoretic tools also yield responses to a whole swath of other objections. I mentioned some other complaints against descriptivism in the last chapter. Detractors object to how fragile the contents of attitudes are according to a descriptivist story. They claim that attitudes are the sort of thing that agents often share and that there are many commonplace phenomena that appear to require that different agents and one agent at different times have beliefs with the same content. The worry is that the descriptivist makes attitude content too fragile and idiosyncratic. I will respond to objections along these lines in sections 6.5 through 6.10 inclusive.



#### 6.4. Regimenting language

Before I proceed, it will be helpful to regiment some language. In the last chapter we concerned ourselves with belief individuation. That is to say, we concerned ourselves with questions such as is it one thing to believe X and another to believe Y or does an agent, by believing X thereby believe Y? What was meant was not that X and Y are the same *token* belief but that X and Y are the same *type* of belief. Let's say that two propositional attitudes X and Y are of the same *type* if, and only if, by having Y an agent thereby has X and by having X an agent thereby has Y.

Philosophers disagree about just how attitudes should be typed, and thereby disagree about when, by having one attitude an agent thereby has another. Some claim that attitudes should be typed by their content (what is believed, feared, desired, etc.). For example, the naïve Russellian claims that two token singular attitudes are of the same type just in case they have the same content, that is, they express the same Russellian proposition.

Others resist this way of typing attitudes, claiming that two attitudes with the same content can be of different types if the content is believed, feared or desired in different ways. Sainsbury and Tye, for instance, claim that beliefs should be typed by their content *and* the concepts involved in the belief.

I am going to talk as if attitudes are typed by their content. If the reader prefers understanding attitudes as typed by their contents plus some other thing, then they ought to replace my uses of the word 'content' below with 'content plus ...' where the ellipsis is filled in with whatever else is supposed to type attitudes. The discussion will proceed just the same.

#### 6.5. Keeping the putative target stable across time and across agents

Suppose a detective named Dirk is investigating the murder of a man named Gordon. At the beginning of the investigation ( $T_1$ ), Dirk has very few beliefs concerning Gordon's murderer. He believes that the murderer did not have any accomplices, and is tall. As the investigation progresses, Dirk comes to believe many more things about the murderer. He comes to believe that the murderer is tall, is somewhat confused about the human circulatory system, etc. Near the end of the investigation ( $T_2$ ), Dirk has come to have all sorts of beliefs concerning Gordon's murderer. By the by, in this case Gordon really was murdered.

It looks as though Dirk's beliefs have a common focus; they are about Gordon's murderer. This sort of case might seem problematic for the descriptivist, as the properties that Dirk associates with the murderer change through the investigation. What features in the content of the belief is not Gordon's murderer itself but the properties Dirk associates with the murderer. For the descriptivist, there does not seem to be an element of the content of Dirk's beliefs that corresponds to Gordon's murderer that shows up in the content of both his  $T_1$  and  $T_2$  beliefs concerning Gordon's murderer, so how can we make sense of the claim that these two beliefs have a common focus?

My answer is that even though Dirk's conception of the putative target changes, the correct test for his beliefs having a common focus is based on the triangulation theoretic schema T3, repeated here for ease of reference.

(T3) For all token attitudes A and B, A is  $g$ -related $_O^S$  to B if, and only if, B satisfies at least one of the triangulation conditions S lays out for O.

Assuming that it is Dirk's standards on triangulation that matter in this case, we can make the appropriate substitutions into T3 and arrive at the following bi-conditional: Dirk's belief at  $T_2$  is  $g$ -related $_{\text{Gordon's murderer}}^{\text{Dirk}}$  to Dirk's belief at  $T_1$  if, and only if, Dirk's belief at  $T_2$  satisfies at least one of the triangulation conditions Dirk lays out for Gordon's murderer, the putative target of his belief at  $T_1$ . Despite there being no element of the content of the two beliefs that corresponds to Gordon's murderer, we can say that both beliefs have a common focus, Gordon's murderer.

In interpersonal cases, the content of two agents' attitudes may not have any element in common that corresponds to the putative target. Suppose Trillian believes that Ford is restless and Zaphod believes that Ford is unhappy. Let us also suppose that Trillian and Zaphod also associate different properties with Ford, the putative target of their respective beliefs. Again, we can test for  $g$ -relatedness by making the appropriate substitutions into T3 and that test will tell us whether the attitudes in question have a common focus.

This way we can claim that *what attitudes are about* is stable across agents and across time, even if there is no element common to the content of both attitudes that corresponds to Ford.

## 6.6. Disagreement and mind-changing

Suppose that Trillian and Zaphod disagree about Ford; Zaphod believes that Ford is happy and Trillian believes that he is not happy. It is possible for agents to disagree about a putative target even when they associate different properties with that target. Fodor and Lepore (2002, 11-14) and others present cases like this as making trouble for theories according to which the contents of attitudes are extremely fragile. If there is no element corresponding to Ford that features in the content of both beliefs, how can we make sense of the fact that they are disagreeing?

Predictably, I propose that the descriptivist ought to make sense of this sort of disagreement by appealing to the triangulation theoretic account of subject matter. To check for disagreement in this case we first have to check whether the attitudes in question are g-related; that is, check to see if they have Ford as a common focus. Then we have to check to see if the beliefs in question ascribe conflicting properties. If both of these tests come up positive, then Trillian and Zaphod are disagreeing about Ford.

A similar case involves an agent changing their mind about the putative target of their attitudes. Suppose at the beginning of the investigation ( $T_1$ ), Dirk believes that Gordon's murderer is careful, whereas near the end of the investigation Dirk comes to believe that Gordon's murderer is not careful. His belief at  $T_1$  and his belief at  $T_2$  conflict but they nonetheless may have the same focus, even by Dirk's own standards. If these two beliefs are g-related in the right way, we can say that Dirk has changed his mind about Gordon's murderer, even if there is no element common to the content of both beliefs that corresponds to Gordon's murderer.

If my suggestion is right, disagreement and mind-changing of this sort should be understood in terms of a common subject matter and the ascription of conflicting properties, *not* in terms of there being a common element in the content of both attitudes and the ascription of conflicting properties.

### 6.7. Agreement and the stability of complete attitudes

We have a test for whether or not attitudes focus on the same putative target, where that target is an object, but there are elements of beliefs that seem to be about properties. If we want to talk about whole attitudes being stable across time we need a way of tracking when two attitudes involve, in some sense, the same property. In this section I will talk about how we might test for

attitudes involving the same property in their subject matter.

Suppose that Trillian and Zaphod both believe that Ford is drunk. If the content of their beliefs depends on what properties they associate with Ford, these beliefs are likely not to have the same content. But it looks like there is a sense in which they believe the same thing. So how should the descriptivist make sense of their ‘believing the same thing’? Again, the triangulation theory supplies the answer. I suggest that the way we should understand ‘believing the same thing’ is as having a belief with the same subject matter, rather than in terms of having a belief with the same content.

We are looking for a test for whether Trillian and Zaphod believe the same thing, in virtue of both believing that Ford is drunk. It looks like there are two things we need to check for: that their beliefs focus on Ford and that the subject matter of their beliefs involves drunkenness.

It should be clear how we can use triangulation theoretic tools to check if the two beliefs concern Ford. To check for this we have to check whether the beliefs are g-related with respect to Ford just as we did in the last section.

The second thing we need to check for is whether drunkenness is in the subject matter of the two beliefs. What does it take for the property ‘drunkenness’ to feature in the subject matter of two attitudes? I propose that we, again, run a triangulation theoretic test but this time for g-relatedness with respect to drunkenness. As I mentioned in section 4.7.4, agents often lay down triangulation conditions for properties. For instance, Trillian presumably believes that there is this property, drunkenness, out in the world, and that there are conditions under which an attitude ascribes or otherwise involves this property. So, making the appropriate substitutions into T3, and relativising to Trillian’s triangulation conditions for the sake of illustration, we arrive at the following biconditional; Trillian’s belief is g-related<sub>Drunkenness</sub><sup>Trillian</sup> to Zaphod’s belief if, and only if, Zaphod’s belief satisfies at least one of the triangulation conditions Trillian lays out for drunkenness. In this way, the property ascribed behaves like a putative quasi-target of the attitudes in question; it forms part of the subject matter of the belief. Call this route the *property triangulation route*.

One might interject as follows: ‘Hang on, according to the descriptivist, properties feature in the content of intentional attitudes. But now the suggestion is that to check for whether or not two beliefs ascribe the same property, we ought not to look to see if the content of the beliefs have a

property ‘drunkenness’ in common, instead we are told to check for a kind of g-relation. If the properties themselves feature in the content of the attitudes, why not just look for the property of drunkenness among the properties in the descriptive content of the belief, when checking to see if the two beliefs ascribe drunkenness?’

My answer is that I would prefer not to *insist* that the property that features in the subject matter is identical to the property that features in the content of both. Sometimes it seems as if two attitudes ascribe what we would like to call ‘the same property’ even when the agents in question have different conceptions of the property in question. As we saw in section 4.7.4, sometimes the main subject matter of a disagreement is a property or set of properties. In cases like these what is at issue is the nature of the property in question. What is more, the parties to the disagreement associate different properties *with the properties*. Yet it still appears that these beliefs are *about* the same property. Even though the agents disagree about what having that property consists in.

Similarly in the Zaphod and Trillian case, suppose Zaphod and Trillian disagree about exactly what it takes to be drunk. There are some borderline cases about which they make different judgments concerning the presence of drunkenness. Suppose that they are aware of this disagreement. But, in this case, Ford is not a borderline case, he is drunk by both of their standards for drunkenness. It seems as if it is not the exact same property of drunkenness that features in the content of both of their beliefs, but it may still be that they ascribe the same property to Ford in the required sense. Just as agents can pick out an object while conceiving of it differently, I wish to leave open the possibility of their ascribing the same property even though they conceive of that property differently.

This brings out another point. Recall the case of Harvey that I discussed in chapter five. Harvey believes that the property of being a lawyer is the property of being a lawyer, but he appears to believe that the property of being a lawyer is the property of being an attorney. He appears to have one of these beliefs and not the other, so there is pressure to say that there is a difference in content. But I want to leave it open that, by Harvey’s own standards for attitudes concerning the same property, these beliefs have the exact same property at their focus, he just doesn’t realise that they do. This claim is unavailable if I insist that for two attitudes to involve the same property in their subject matter, this property must feature in the *content* of both of the attitudes. So I would rather apply the triangulation theoretic test for a property featuring in the subject matter.

That being said, I am not ruling out the possibility that *sometimes* the property that features in the subject matter of both attitudes also shows up in the content of both attitudes. If the reader would prefer to only use the triangulation theoretic tools in the context of object targets, it is open to them to formulate a different test for attitudes ascribing the same property. Perhaps the property needs to somehow feature in the content of both attitudes. I have given my reasons for preferring the more full blooded triangulation theoretic approach, but most of the discussion below will proceed in the same way whether we take the property triangulation route I prefer or adopt some other test for a property being the in the subject matter of attitudes.

So we have a test for whether Trillian and Zaphod's beliefs are g-related by having Ford as a common focus and we have a test for g-relatedness with respect to drunkenness. If both these tests come up positive, and the two beliefs are g-related relative to Ford and the property of being drunk, then, by both believing that Ford is Drunk, Trillian and Zaphod 'believe the same thing'.

We can make a similar move in intra-personal cases. Suppose at  $T_1$  Dirk believed that Gordon's murderer is tall and later on in the investigation at  $T_2$  he still believes that Gordon's murderer is tall. By descriptivist lights, these two beliefs may have different contents, but there seems to be a sense in which Dirk believes the same thing at  $T_1$  and  $T_2$ . Again, the idea is to run the sameness of subject matter test on the putative target, Gordon's murderer, *and* the property ascribed. If it turns out that the two beliefs stand in the appropriate g-relations, then it is correct to say that Dirk believes the same thing at  $T_1$  and  $T_2$ .

In this way triangulation theoretic tools can provide an account of the sense in which different agents, and the same agent at different times believe the same thing, fear the same thing, desire the same thing, etc., *without* being committed to their having attitudes with the same content. The usual way in which 'believing the same thing' is understood is in terms of having a belief with the very same content. Since it is a feature of my preferred descriptivist theory that belief content is extremely fine-grained, anyone who is tempted by my preferred kind of descriptivism should not adopt this way of thinking. If the above suggestion is right, they should instead understand the sense in which different agents 'believe the same thing' in terms of subject matter. Agents 'believe the same thing' just in case they have beliefs that are g-related relative to all the targets of those beliefs and all the properties they ascribe. Similarly with other attitudes, we can apply the same strategy to 'fearing the same thing', 'desiring the same thing', and so on.

On this account, agents will *often* believe, fear or desire the same thing. The content of attitudes may be extremely fragile, but subject matter is often much more robust.

## 6.8. Communication

Next, consider the theory of successful communication that Stalnaker (1978, 1988, 151) defends, and Weber (2013, 207-208) calls *the FedEx model of communication*. The FedEx model says that successful communication involves an agent expressing a belief and their audience coming to have the same belief. Suppose Trillian believes that Ford is ill and wants to tell Zaphod about it, so she utters the sentence ‘Ford is ill.’ According to the FedEx model, the communicative act is successful if Zaphod comes to believe the same thing that Trillian believes, that Ford is ill. According to the FedEx model, the belief that the audience comes to believe is supposed to be identical to the belief expressed by the speaker, at least in cases of successful communication.

It is often supposed that for communication to be successful, the belief that the speaker expresses and the belief that the audience comes to have are supposed to have the same *content*. This is certainly how Stalnaker himself conceives of things. But if the descriptivist is right, this is too high a bar to set for successful communication, since agents do not often have attitudes with the same content. It looks as though there is often successful communication in cases in which the beliefs in question have *different* content.

The descriptivist ought to make the same sort of move that we made above with regard to agreement and ‘believing the same thing’ in section 6.5. There is a rival of the FedEx view that says when communication is successful, the audience comes to have a belief that is suitably related to the attitude expressed by the speaker. I have in mind a model according to which the communicative act is successful, as long as the audience comes to believe the same thing as the speaker, in the sense discussed in the previous section. For example, when Trillian believes that Ford is ill and expresses that belief by uttering ‘Ford is ill’ and Zaphod comes to have a belief, we can say that the communicative act is successful, by Trillian’s standards, if Zaphod’s newly formed belief is g-related<sub>Ford</sub><sup>Trillian</sup> and g-related<sub>Illness</sub><sup>Trillian</sup> to Trillian’s.

## 6.9. Other features of my approach

At this stage, let me say a few more things about the general approach I am defending that will hopefully both distinguish it from other approaches and provide reasons to find my approach attractive.

The package of views I have been defending does not involve appealing to any constraints on the content of intentional attitudes that problematically *gloss over* the details of how the agent takes things to be. In chapter five I discussed some of the benefits and costs of appealing to external constraints on what attitudes agents have. Theories like descriptivism, according to which attitude contents are extremely fine-grained are often criticised because they make it too hard for agents to have attitudes with the same content or to have the same target in mind. In response to this concern, many seek to keep attitudes stable by appealing to worldly facts such as causal chains, facts about the agent's community or objective eligibility facts (see chapter five section 5.2). If constraints of one or more of these sorts are in play, some differences in how the agents take the putative target to be and differences in their behavioural dispositions are *glossed over*, that is, small differences in the agent's mental state will not necessarily make a difference to what attitudes they have. Suppose Trillian and Zaphod both believe that Ford is drunk and they associate different properties with Ford such that if all we had to go on when assigning content to those attitudes was the associated properties, then their beliefs would *not* have the same content. If there are external constraints in play, their attitudes may nevertheless have the same content. The world helps attitudes get their content and, for this reason, it is easier for attitudes to have the same content in spite of differences between the respective agents' (narrowly construed) psychological states.

There are at least two problems with appealing to external constraints on content in this way. The first is that glossing over these differences gives rise to conflation, see sections 5.2 and 5.4. The details that those who adopt this anchoring strategy gloss over *do* appear to make a difference to the content of attitudes. By glossing over these details they give rise to conflation; there is pressure to think that it is one thing to believe X and another thing to believe Y but, according to the offending theories, to believe X *just is* to believe Y.

The second is that these worldly resources are not always available and all the phenomena that are usually linked to shared content occur in spite of their being unavailable. Agents can agree, disagree, communicate, and keep the target of their attitudes fixed across time even when the relevant attitudes are empty and relevant external anchors are missing. When determining the contents of



empty attitudes, for instance, there are no causal links between the object that the attitude is about and either the believer or members of the believer's community. Either this is because the object in question is understood to be abstract, merely possible or non-existent, or because the object is missing entirely. So there are cases in which we are forced to analyse the phenomena traditionally explained by common content without appealing to worldly constraints of this sort. But once we have explained these phenomena *without* appealing to these worldly constraints, what is stopping us from applying that kind of story to all cases? Even the cases in which those worldly resources *are* available. There appears to be no reason not to.

What is more, there are at least two reasons to avoid appealing to external constraints altogether. The first is uniformity. All else being equal, we ought to prefer uniform theories of communication, agreement, disagreement and 'believing the same thing'. Given that uniformity is a virtue and that we have to make sense of these phenomena without appealing to external constraints on content in *some* cases, the cases in which there are no such constraints available, there is pressure to explain these phenomena in the same way even in cases in which one *could* appeal to extra constraints.

The second reason arises from the costs of appealing to external constraints on the content of attitudes. As I argued in chapter five, appealing to external constraints gives rise to conflation, which is the inability to make certain distinctions between attitudes. This makes the appeal to external constraints somewhat costly. If we apply a theory that does not require external constraints on content across the board, we can avoid this cost altogether.

But what about the reasons why the external constraints were deemed necessary in the first place? What about the reasons that make appealing to these constraints useful and maybe worth the costs? Many of these reasons were discharged in sections 6.5 to 6.8 inclusive. If the descriptivist helps herself to the triangulation theoretic resources, she can make sense of how subject matter stays stable across agents and across time, and account for the phenomenon of agents believing or fearing the same thing, all without appealing to external constraints on the content of those attitudes.

Descriptivism and the triangulation theory fit together in another way. According to both theories we derive guidance about the properties of the intentional attitudes from the details of how the agent takes the putative target of their beliefs to be, rather than from the putative target itself or from the relationship between the holder of the attitude and the target object. If descriptivism is

right, then how the putative target is taken to be by the believer makes a difference to the *content* of intentional attitudes. If the triangulation theory is right, how the putative target of the belief is taken to be also makes a difference to how the *subject matter* of attitudes ought to be understood, at least when the triangulation conditions of the attitude's holder.

It is true that these moves involve reinterpreting the phenomena somewhat. For instance, the intuitive conception of 'believing the same thing' as 'having a belief with the same content' will have to be rejected. The descriptivist will have to resist this natural way of understanding the phenomena, but the benefits of the descriptivist theory outweigh this intuitive cost of the view.

There is another cost of the package of views I defend. Since my approach involves using triangulation theoretic resources to make sense of agreement, disagreement, communication, and the stability of attitudes across agents and across time, our account of these phenomena will end up being relative to which triangulation conditions are in play. We will not be able to say which agents agree, disagree or communicate simpliciter. We will only be able to say that two agents agree, disagree or communicate *relative to a particular set of standards*. For example, it may be that by one set of standards Trillian and Zaphod disagree, but by some other set of standards they do not disagree, since according to one set of standards, but not the other, their attitudes have a common focus. One might have hoped for a non-relative, once and for all, theory of these phenomena. My approach cannot deliver this. No doubt this is a cost of my approach, but it is a cost that we should be willing to pay.

#### 6.10. Uniformity again

Because of the threat of conflation, the descriptivist has an advantage over many of their rivals who appeal to causal, eligibility or community facts to anchor down the content of attitudes. They have this advantage even if we only consider cases in which those resources are available. There is also some pressure to tell a uniform story about the content of empty and non-empty attitudes.

Even if other theories handled the content of empty intentional attitudes as well as the descriptivist story does, there would still be reason to prefer the descriptivist theory in the empty case. The descriptivist does a better job at accounting for the content of non-empty intentional attitudes, since it does not lead to the problems discussed in sections 5.2 and 5.4, and there is pressure to tell a

uniform story so we can push for descriptivism across the board.

So, in a sense, the desire for uniformity pushes for a descriptivist theory of mental content *in both directions*. Descriptivism provides an attractive account of empty intentional attitudes and since uniformity is a virtue, there is pressure to adopt a descriptivist account of non-empty attitudes as well. In the other direction, the descriptivist theory allows us to avoid conflation, by making the appropriate fine-grained distinctions between non-empty attitudes. So, again, since uniformity is a virtue, there is pressure to adopt a descriptivist picture of content of empty attitudes.

#### 6.11. Traditional arguments against the description theory of names

So far in this thesis I have focused primarily on issues in the philosophy of mind, addressing questions like: ‘how should we understand the subject matter of intentional attitudes?’, ‘when do two attitudes have the same subject matter?’, and ‘how should we understand the content of intentional attitudes?’ But many reject the descriptivist theory of the content of attitudes because they reject its cousin, the description theory of proper names. There are well known arguments that are supposed to show that the description theory of names is inadequate and many, I suspect, reject the descriptivist theory of the content of attitudes because they think similar arguments knock out this account as well.

I will now present some arguments that have been levelled against description theory of names, explain how these arguments might be relocated and levelled against the descriptivist theory of the content of attitudes. I will then argue that these relocated arguments can be answered and are not fatal to a descriptivist theory of the content of attitudes.

According to the description theory of proper names, originally proposed by Russell (1905, 1910-1911), proper names are, in fact, disguised descriptions of the form ‘the such and such’ and ordinary proper names refer in just the same way descriptions do. On this sort of view, the semantic value of a name is identical to the semantic value of some description or other, associated with the name, and the name *denotes* or *refers to* whatever satisfies the description. I am *not* defending the

description theory of names here.<sup>31</sup> My goal is to investigate whether any of the arguments that are often taken to be fatal to the description theory of names can be relocated and effectively levelled against the descriptivist theory of mental content.

#### 6.11.1. The argument from ignorance and error

The first argument against the description theory of names I will discuss is, perhaps, the most influential. It is sometimes called *the argument from ignorance and error*, or *the semantic argument*. This argument was presented forcefully by Donnellan (1970, 353-354) and Kripke (1980) but examples like these crop up all over the literature. I will use Kripke's case involving the name 'Gödel' to illustrate the objection. The argument goes something like this. Suppose after taking a brief course in mathematics, I believe that Gödel proved the incompleteness of arithmetic. I associate the description 'the person who proved the incompleteness of arithmetic' with Gödel. Suppose also that the following story is true: The man who is often called 'Gödel' did not, in fact, prove the incompleteness of arithmetic. A man named Schmidt who was found dead under mysterious circumstances actually did the work. Gödel somehow got hold of the manuscript and the proof was, from that point on, attributed to Gödel.

Kripke's objection to the description theory of names is that even though it is Schmidt that best fits the description I associate with the name 'Gödel', when I use the name I refer to Gödel and not Schmidt. This is supposed to show that reference does not go by the description I associate with the name, but by some causal chain of communication leading back from my use of the name to 'Gödel', the man himself.

It will be helpful to outline one kind of reply defenders of the description theory of names, such as Loar (1976), Kroon (1987, 2009), Lewis (1984, 1997), and Jackson (1998, 2010), often give to this argument. The reply goes like this: Kripke is right about what the name refers to in the case described, but wrong to think that this is a problem for the description theory of names. What cases like these show is that often the descriptions associated with names are not the descriptions that

---

<sup>31</sup> I am also sympathetic to the description theory of names. I have been convinced mainly by the work of Kroon (1987, 2009) and Jackson (1998, 2010).

first come to mind, involving famous deeds of the putative target, but a more complicated description involving causal or meta-linguistic properties. Sometimes the descriptions associated with names are causal (e.g. ‘the thing at the back of the causal chain leading back from my use of the name “Gödel”’), and other times they are meta-linguistic (e.g. ‘the thing that experts in my community pick out when they use the name “Gödel”’). If one of these descriptions is the one I associate with the name “Gödel”, then the description theory of proper names can deliver the desired result about the Gödel/Schmidt case and cases like it. Incidentally, I think this response is satisfactory and that the description theory of proper names survives this attack.

It is also worth noting we *might* have associated the description ‘the person who proved the incompleteness of arithmetic’ with ‘Gödel’. Kripke’s case shows that we probably didn’t but that does not mean that we could not have. As Kripke (1980, 91) admits, there is nothing stopping an agent from using the term in that way.

Can this kind of argument be transformed into an argument against the descriptivist theory of the content of attitudes? Suppose I associate the properties of being the prover of the incompleteness of arithmetic with Gödel, the putative target of my belief. Suppose that the story about Gödel and Schmidt told above is true. It seems as if, in this scenario, we want to say that my belief is about Gödel and not Schmidt. So if the descriptivist were committed to the claim that my belief is about Schmidt in such a case, she would be in trouble.

But there is no reason whatever to suppose that the descriptivist will deliver this result. The descriptivist ought to claim that *those* associated properties do not determine what my belief is about in this case. Rather, it is the association of the putative target with certain causal and metalinguistic properties that determines what the belief is about. For instance, I might associate the description ‘the thing I heard of under the name of Gödel’ with the name ‘Gödel’. If these are the associated properties that determine what the beliefs are about, then the descriptivist can deliver the correct verdict that the belief is about Gödel, even though I am wrong about him. One limitation of this response is that it depends on a distinction between the associated properties that determine what the belief is about and those that do not.

However, if the descriptivist helps themselves to triangulation theoretic resources, she will have a good way of drawing this distinction. Often agents believe that the putative target of an attitude is

the sort of thing about which agents can be mistaken. I might, for instance believe that you and I disagree about an object *without* thinking that we must be thinking about different things, in virtue of our having different beliefs about the target. Suppose in the Gödel/Schmidt case a friend and I disagree about whether Gödel proved the incompleteness of arithmetic. It may be that, by my lights, we are disagreeing about Gödel (both our beliefs are about Gödel) even though at least one of us must be mistaken about him.

This will be reflected in the triangulation conditions I lay out for the putative target, Gödel. If what a belief is *about* is understood in a triangulation-theoretic way, we can say that associated properties *do not* determine what a belief is about just in case, by the standards of sameness of subject matter individuation in play, one can be mistaken about whether the putative target has that property.

There is another way we might try to relocate the argument from ignorance and error into the realm of mental content. The relocated argument I have in mind concerns the content of attitudes rather than what they are about. The argument might go something like this: ‘agents associate all sorts of properties with targets of their beliefs. But if all of those properties feature in the content of beliefs concerning the target, we run into trouble. If the agent is wrong about that object having *any* of those properties, *all* the beliefs concerning that putative target will come out false. Suppose I believe that my mother’s favourite colour is orange and associate the property of being a lover of the colour orange with her. Unfortunately, I am wrong, her favourite colour is green and she doesn’t like orange at all. When I believe that my mother is happy the descriptivist will say that the content of my belief is for some *x*, *x* is a lover of the colour orange and (has all the rest of the properties I associate with my mother) and happy. But this belief will be false *because I am wrong about her preference for orange* and this is surely the wrong result. So there is a problem with the descriptivist theory of content.’

I want respond to this objection with a move that is analogous to the move that causal descriptivists make in the context of the description theory of names. What cases of this sort show is that often only a certain subset of the properties an agent associates with the putative target of their beliefs feature in the content of singular beliefs they have about that putative target. In most cases they are the kind of casual, metalinguistic properties discussed above. For example, the content of my belief that Gödel proved the incompleteness of arithmetic might be something like ‘for some object

x, x is the object at the back of such and such causal chain and x proved the incompleteness of arithmetic'. In this way many other mistaken beliefs I might have about Gödel do not threaten to render *this* belief false. On this sort of story, often only a certain set of the properties an agent associates with a target feature in the content of beliefs directed at that target. Which ones are given this special status? Which of the associated properties are given special status will reflect their beliefs about the target, the kind of thing it is and its place in the world.

### 6.11.2. The epistemic argument

Another argument against the description theory is what is sometimes called the *epistemic argument*. The argument goes like this: if the name 'Aristotle' is equivalent to a description like 'the teacher of Alexander the Great', then the sentence 'Aristotle, if he or she exists, taught Alexander the Great' should be true, and we would be able to know that it is true a priori since it is equivalent to the sentence 'the teacher of Alexander the Great, if he or she exists, taught Alexander the Great'. But the fact that Aristotle taught Alexander the Great is not knowable a priori. To discover that fact we had to look at history books and what not. So, the description theory of names is false since it leads to false claims about the epistemic status of sentences. We can, it seems, run this same argument for any description that mentions a fact about the putative target that is known a posteriori.

Is there a similar argument that can be levelled against the descriptivist theory of the content of attitudes? Suppose the descriptivist theory is true and I associate the property of being the teacher of Alexander with the putative target Aristotle. Does it follow that I can be certain that my belief that Aristotle taught Alexander is true simply in virtue of how I frame my beliefs concerning Aristotle? No. As I mentioned before, I am likely to think that Aristotle is the kind of object about which I can be mistaken, and in particular, mistaken about whom he taught. My beliefs about the putative target and his place in the world lead me to believe that I can be wrong about the answer to the empirical question of whom he taught.

### 6.11.3. The modal argument

There is another argument that is often presented against the description theory of names called *the modal argument*, often associated with Kripke (1980, 30). The argument goes something like

this: according to the description theory of names, the sentence ‘Aristotle might not have taught Alexander’ is equivalent to some sentence like ‘The teacher of Alexander might not have taught Alexander’. (In these sentences ‘might’ is supposed to be given its metaphysical reading rather than its epistemic reading.) However, says the objector, the first sentence is true while the second is false. So the description theory of names leads us to incorrect verdicts about the truth values of certain modal sentences.

Is there an analogous argument that undermines the descriptivist theory of attitude contents? For instance, can my belief that Aristotle might (metaphysical reading) not have taught Alexander be true *even if* I associate the property of having taught Alexander with Aristotle? The descriptivist should respond in much the same way that Dummett (1973, 110-151) responded to the modal argument on behalf of the description theory of names. When evaluating the content of modal beliefs, the descriptivist needs to be careful about which modal property is being ascribed to which putative target.

The descriptivist need only say that the belief that Aristotle might not have taught Alexander is equivalent to the belief that the thing that has the properties the believer associates with Aristotle (which, *ex hypothesi*, includes the property of having taught Alexander) might not have taught Alexander. This belief may be true and consistent with the belief that Aristotle taught Alexander. There is no harm as long as we do not confuse this belief with another belief that is plainly false; the belief that there might have been some object *x*, such that *x* taught Alexander and *x* did not teach Alexander.

#### 6.11.4. The ‘passing the buck’ argument

Some complain that the description theory of names *passes the buck*. This complaint is pressed forcefully by Devitt (1996, 159). The objection goes like this: according to the description theory of names, the semantic value of names is explained by an appeal to other words, the ones that feature in the relevant description. But how do *those* words get their semantic values? Perhaps the answer is ‘by being equivalent to some other description’ but then the complication simply arises again. Eventually, runs the objection, some words will have get their semantic value in a way that is not parasitic on the semantic value of other words. As Devitt (1996, 159) puts it, ‘[d]escription theories pass the buck. But the buck has to stop somewhere.’



Although this is presented as a worry concerning the description theory of names, I think a similar challenge faces the descriptivist about the content of intentional attitudes. But before I discuss the relocated problem it will be instructive to consider the moves available to the defender of the description theory of names when faced with Devitt's challenge.

Russell (1910-1911, 117) has an answer to this objection. For Russell, the components that form part of a description are properties with which we are acquainted. I can understand a description just in case it is composed wholly of constituents with which I am acquainted. To take an example, I can understand the description 'the only red ball in this bag' because I am acquainted with the property *redness* (picked out by the predicate 'red') and I know what it is to be in 'this bag', presumably by description, there will be some other description that is composed of elements with which I am acquainted. On this kind of view, there is a special class of words that have their semantic value in a way that is not parasitic on the semantic value of other words. The special words are those with whose semantic values I am acquainted. If this is right, I am acquainted with a bunch of properties, the semantic values of some special predicates, and I build up descriptions from those raw materials. Russell's view, the details of which involve a commitment to *sense data*, has fallen on hard times. But quite apart from Russell's particular implementation of this strategy, there is an approach, friendly to the description theory of names, according to which there are certain special words or concepts that have non-parasitic semantic values and all the others get their semantic values *built up* out of these special words or concepts.

Another option open to a defender of a description theory of names is to adopt a holistic account of the semantic value of language. They might claim that all the terms in a language get their semantic value at once. According to this view, the semantic value of each element of the language depends on its place in the whole language; every piece of the language is parasitic on every other piece of the language. On this sort of story, the buck stops at the whole language.

There are many complaints made about holistic approaches to meaning,<sup>32</sup> but one of them is particularly relevant for present purposes. Fodor and Lepore (2002, 11-14) worry that, since the semantic value of words depends on an agent's whole theory, small changes in the whole theory will

---

<sup>32</sup> For a polemical but interesting discussion of the merits and costs of holism see Fodor and LePore (1992, 2002).

change the meaning of the constituents. But this makes the meaning of a particular term unduly fragile. We will have a hard time explaining the sense in which the meaning of terms stays constant across time and across agents. Since almost every agent has a different complete theory, it will be unduly hard for two agents to mean the same thing by the words they use. Stalnaker (2003, 212) makes a similar complaint against holistic semantic theories.

There is a challenge very similar to the passing the buck challenge facing the descriptivist theory of attitude contents. The descriptivist says that it is the properties associated with the putative target that feature in the content of attitudes and not the object itself. But how do the believer's representations of those properties get their content? The answer might be 'in terms of some other associated properties' but then the complication arises again. Eventually some elements of the content of attitudes will have to have their content in a way that is not parasitic on the content of other elements of content. What can the descriptivist say in response to this challenge? I will discuss three responses.

The first response is unsatisfying and philosophically uninteresting, but it ought to be dialectically effective. The descriptivist might respond that this problem is a problem for all sides of the debate and, for this reason, is not a reason to reject descriptivism. *Everyone* needs a story about how agents come to represent properties, even the most avid naïve Russellian. It is true that there are puzzles concerning how this representation goes, but since this challenge faces almost all theories of the content of attitudes it is hardly fair to pin this on the descriptivist. This response is reasonable but leaves much to be desired. A problem for everyone is not a problem for no one, despite the temptation that some feel to make this claim when fending off objections.

Another option open to the descriptivist is to claim that some elements that make up the content of attitudes get their content *by acquaintance*. I am imagining a move similar to the move made by Russell in the case of linguistic content that relies on agents being acquainted with properties in the world. There might be a certain special set of foundational concepts that an agent possesses in virtue of being acquainted with things in the world in the right sort of way. These concepts might then be used as building blocks out of which other components of content are built. Call this the *foundationalist* brand of descriptivism. I take this response to be a kind of promissory note. Whether or not this move is a good one will depend on our finding an adequate theory of acquaintance. An interesting avenue of further research would involve exploring ways of developing a

foundationalist approach to mental content.

Another option is to adopt a holistic picture of the content of attitudes. This move is structurally similar to the move to a linguistic holism discussed above. The idea is that each element of the content of attitudes gets its content from its place in the agent's total mental configuration. The buck stops at the whole network of intentional attitudes. Call this the *holistic* brand of descriptivism.

One major complication with this sort of mental content holism is that it is open to a complaint analogous to the fragility complaint discussed above. Complete mental configurations are often different from moment to moment and from agent to agent, and if holism is true *every* such change will mean a change in the content of all the individual parts of the whole. This makes attitudes overly fine-grained in a worrying way; agents cannot believe the same thing, and disagreement looks hard to explain, and so on. We are back at the now familiar objections to fine-grained accounts of attitude content discussed in sections 6.4 to 6.7 inclusive.

Predictably, I suggest that the descriptivist about mental content who is also tempted by the holism ought to respond by appealing to triangulation theoretic resources. In fact, if someone is independently tempted by mental content holism, the triangulation theory can provide a response to the instability complaint, whether or not they are also attracted to a descriptivist theory of the content of attitudes. The *content* may be fine-grained and fragile, but this need not worry us. We can make sense of the phenomena usually linked to identity of content by helping ourselves to a triangulation theoretic conception of subject matter as outlined above.

There is another challenge that faces the holistic brand of descriptivism. Where does the whole get its content? Lewis (1974) has some interesting things to say in answer to this sort of question. Roughly, Lewis's idea is that the whole gets its content from its total *functional* profile. The question of whether this sort of approach can be made to work is beyond the scope of this thesis, let alone this chapter. The question of how total mental states can get their content is a standing challenge for any holistic view of mental content and Lewis's idea is just one way that the holist might explain how the total mental state gets its content. The triangulation theory does not yield an answer to *this* challenge. But the triangulation theory does allow us to live with any fragility of content that might stem from a holist conception of mental content.

## 6.12. Conclusion

In this chapter I have argued for two claims. Firstly, I have shown that if we adopt a particular theory of subject matter and put it to work in a particular way, we can explain phenomena that are usually associated with attitudes having the same content without requiring that agents often have attitudes with the same content. In this way we can gain all the advantages of fine-grained attitudes while avoiding many of the problems that are traditionally pinned on accounts with this feature. Secondly, I have argued that some of the most common arguments against the description theory of names cannot be transformed into arguments that effectively undermine the descriptivist theory of the content of attitudes, at least when this theory is supplemented by a triangulation theoretic account of subject matter.

## CHAPTER 7: A SOLUTION TO KRIPKE'S PUZZLE ABOUT BELIEF

### 7.1. Is puzzling Pierre inconsistent?

In this chapter I will use triangulation theoretic tools to solve a puzzle about belief due to Kripke. Kripke (1979) illustrates his puzzle with an example involving a Frenchman named Pierre. The story I am about to tell is very much like the one Kripke presents, but I will change things slightly to avoid distracting and irrelevant details.

Suppose there is a Frenchman named Pierre. When he was growing up in France he heard of a famous foreign city that people called 'Londres'. He heard many interesting things about this city and, on the basis of what he heard, came to believe that Londres is pretty. When Pierre reaches adulthood he leaves France and travels the world and eventually ends up living in London. Unfortunately, he winds up in an ugly part of London. The people with which he interacts in London call the city in which Pierre now lives 'London' and Pierre adopts this name for his new home. He comes to believe that London is not pretty. He never realizes that the city in which he now lives is Londres, the city he heard of when he was growing up in France. Pierre also holds onto the belief that he formed in his childhood that Londres is pretty. As it turns out, Londres *just is* London; they are the same city.

The description of the case mentions two of Pierre's beliefs, his belief that Londres is pretty and his belief that London is not pretty. The puzzle concerns if and how these beliefs conflict. (Notice that according to this presentation of the puzzle, it is a puzzle about beliefs and not primarily a puzzle about belief *reports*. Kripke's paper also raises issues about *translation* and *belief reports* about which I will not have anything to say here.) A solution to the puzzle would amount to a satisfactory answer to the following question: does Pierre have contradictory beliefs concerning London?<sup>33</sup>

---

<sup>33</sup> Given a certain interpretation of 'pretty', Pierre need not have contradictory beliefs about London. London might be pretty in parts and not pretty in parts. Kripke clearly intended us to interpret Pierre's beliefs as ascribing the property of being pretty to London as a whole.

## 7.2. Yes and No

I wish to answer ‘yes and no’. There is a sense in which Pierre has contradictory beliefs concerning London and a sense in which he does not.

Let me tease out the sense in which he does and the sense in which he does not by comparing the original Pierre case to two variations. When I say ‘original Pierre case’ I don’t mean the cases originally presented by Kripke, I mean the case outlined in section one. I will start with the sense in which, in the original case, Pierre *does* have conflicting belief concerning London. Consider a case that is just like the original story of Pierre except that there are two different cities with the names ‘London’ and ‘Londres’. Call this case *LuPierre* (for lucky Pierre). In this case it seems as if Pierre’s beliefs are, somehow, even less conflicting than they were in the original case. In the original case, London and Londres are *in fact* identical but Pierre lacks this information. Intuitively, we might say that Pierre’s beliefs are about the same thing though he does not realize it. The distinction between *LuPierre* and the original case appears to show that the facts about what is identical to what seem to make a difference to whether his beliefs are conflicting. Pierre’s beliefs in the original case conflict in a way that his beliefs in *LuPierre* do not. In the original case, Pierre has conflicting beliefs concerning London *in this sense*.

What about the sense in which he does not have contradictory beliefs about London? To bring this out let us consider a variant on the original Pierre case. Suppose that Pierre is aware that London and Londres are the same city but refuses to revise either of his two beliefs. He has a strange view of the world, according to which cities can have contradictory properties. He believes that London is the same city as Londres but still thinks that it is pretty and not pretty. Call this case *InPierre* (for inconsistent Pierre). I want to say that, in this scenario, Pierre has beliefs concerning London that are inconsistent in a way that they are *not* in the original case. In the original case, Pierre does not have contradictory beliefs concerning London in *that* way. The distinction between *InPierre* and the original case suggests that whether Pierre *realizes* that his beliefs are about the same putative target makes a difference to whether his beliefs are contradictory, at least in some sense.

If the reader dislikes the idea of Pierre having an explicitly inconsistent conception of London, perhaps a more mundane case will serve to illustrate the point. Suppose there is one sentence written on the blackboard in a room. Suppose Freda is sitting in the next room over and believes that

there are two sentences on the blackboard in the first room, one written above the other. Freda believes that the sentence furthest up on the board is true and the sentence furthest down on the board is not true. Compare this to a case just like it except it is a priest sitting in the next room over. The priest believes that the sentence furthest up on the board is true and the sentence furthest down on the blackboard is not true *and* that they are one and the same sentence (he believes the sentence is both true and not true). The priest has inconsistent beliefs concerning the sentence in a way that Freda does not. In this sense, Freda does not believe inconsistent things concerning the sentence in the same way that the priest does.

A good account of the Pierre cases and cases like it will yield an account of the sense in which Pierre has contradictory beliefs about London and the sense in which he does not. The rest of this chapter will be devoted to finding an account of attitudes that fits this bill.

### 7.3. Unsatisfactory accounts

One theory that has trouble making sense of cases like Pierre's is the naïve Russellian theory. In the original Pierre case London and Londres are the same city. Since, for the naïve Russellian, the content of the two beliefs involves the very same object, they can account for the sense in which Pierre *does* have contradictory beliefs; his beliefs ascribe contradictory properties directly to the very same object. The contents of the beliefs are that London, the city itself, is pretty and that London, the city itself, is not pretty.

The problem is that the naïve Russellian is then forced to claim that *in the original case*, Pierre has contradictory beliefs, just like he does in my InPierre case. Intuitively, for the Russellian, what matters is what is, in fact, identical to what. There is no room to allow Pierre's realising or failing to realise some identity fact to make a difference to whether he has contradictory beliefs concerning London. Framing things in my way, we can say that the naïve Russellian theory is unable to appropriately distinguish between Pierre's beliefs in the original case and Pierre's beliefs in InPierre; the naïve Russellian mistakenly treats Pierre's beliefs in the original case as being inconsistent in the same way that they are in InPierre. Salmon (1986, 1989, 2006) and Soames (1987) accept this conclusion. But this account is unsatisfactory. I agree with Lewis (1981) and Kripke (1979, 269-270) that *this* kind of inconsistency should not be attributed to Pierre. Since the naïve Russellian theory implies that it should, the naïve Russellian theory is mistaken, or at least in need

of significant revision.

Next, consider a straightforward descriptivist account of Pierre's beliefs, defended by Lewis (1981). For the descriptivist, the content of Pierre's belief that London is not pretty is that the thing that plays the London role for Pierre is not pretty. Likewise, the content of his belief that Londres is pretty is that the thing that plays the Londres role for Pierre is pretty. This kind of picture captures the difference between Pierre's beliefs in the original case and his beliefs in the InPierre case. In the original case he does not *believe* that the thing that plays the London role is, in fact, the thing that plays the Londres role.

The problem with the flat-footed descriptivist response is that it is not clear how the descriptivist captures the sense in which Pierre *does* have contradictory beliefs concerning London. The object that actually plays either of the roles does not feature in the content of the beliefs, so it seems as if the fact that Londres is the same city as London does not make a difference to the content of Pierre's beliefs. The difference between the original Pierre case and LuPierre revealed that what is, in fact, identical to what, appears to make some difference to whether Pierre's beliefs are inconsistent in some sense. Since, for the descriptivist, London, the object itself, does not feature in the content of either of Pierre's belief, it is not clear how facts about what is identical to London can make any difference to the consistency or inconsistency of his beliefs concerning London. In fact, as I discussed in chapter six, one might worry that the straightforward descriptivist theory, unaugmented by a theory of subject matter, cannot even account for the sense in which either of his beliefs are *about* London in any sense. I should mention that Lewis does have an answer to this complaint which I will discuss presently. This answer involves *augmenting* the straightforward descriptive theory. Without such augmentation, the descriptivist theory struggles to handle the original Pierre case.

Naïve Russellianism allows us to make one of the required distinctions, and descriptivism allows us to make the other. Perhaps we should adopt a view that incorporates the virtues of both.

Lewis's (1986, 33) idea is that, in addition to the descriptivist content of beliefs, agents are often 'acquainted' with objects either perceptually or via longer causal chains of information transmission. In the original case Pierre is acquainted with London in two different ways, he is London-acquainted with London and Londres-acquainted with London. These are two different ways for



beliefs to come to be about the same thing. Remember that the descriptivist can account for the difference between the original case and InPierre but had trouble adequately distinguishing the original case and LuPierre.

With this acquaintance machinery in place, the descriptivist can also adequately distinguish between the original case and LuPierre. In the original case, but *not* LuPierre, Pierre is London-acquainted with an object *x* and Londres acquainted with an object *y* *and*, in fact, *x* is identical to *y*. In LuPierre, Pierre is London-acquainted with an object *x* and Londres-acquainted with an object *y* but, in fact, *x* is not identical to *y*. In the original case London, the city itself, ‘gets into the act’ (1986, 33) because it is the object that Pierre is, in fact, acquainted with. Which objects are identical to which in Pierre’s environment thus make a difference to whether he has conflicting attitudes about London.

A somewhat similar strategy for handling the original Pierre case involves adopting a type of pluralism. Chalmers (2002a) suggests that we take this sort of approach to cases like Pierre’s. For an outline of Chalmers’s two dimensional picture of the content of attitudes, see section 5.6.2. The idea is to claim that Pierre’s beliefs both have two kinds of content, epistemic and subjunctive, and that, in the original Pierre case, they have conflicting subjunctive content but consistent epistemic content.

For the sake of illustration let us assume a descriptivist interpretation of the epistemic intension. (My evaluation of the Chalmers-style treatment of Pierre-type cases will not rely on this interpretation.) The epistemic content of Pierre’s belief that London is not pretty is the set of worlds at which the thing, whatever it is, that has the properties Pierre associates with London is not pretty. The epistemic content of his belief that Londres is pretty would then be captured by the set of worlds in which the thing, whatever it is, that has the properties Pierre associates with Londres is pretty. These contents do not conflict. The sets of worlds that capture the content of these two beliefs overlap.

But, by Chalmers’s lights, the two beliefs have conflicting subjunctive content in the original Pierre case. The subjunctive content of his belief that London is not pretty is the set of worlds in which the thing that *actually* has the properties that Pierre associates with London is not pretty. The subjunctive content of his belief that Londres is pretty is the set of worlds in which the thing

that *actually* has the properties Pierre associates with Londres is pretty. In the standard Pierre case the thing that has the Londres properties *just is* the object that has the London properties. In this way the subjunctive contents of these beliefs are contradictory since there is no possible world at which that object is both pretty and not pretty.

The move to Lewis's acquaintance-based approach or to a Chalmers-style pluralism might seem, at first sight, very satisfactory. As discussed above, someone who adopts Lewis's acquaintance-based approach can capture the relevant difference between the original Pierre case and LuPierre. They can also distinguish the original Pierre case from InPierre; in one, but not the other, Pierre associates the property of being identical to London with Londres, and vice versa.

The Chalmers-style pluralist can capture the difference between the original Pierre case and LuPierre. In the case LuPierre, Pierre's beliefs have consistent epistemic and subjunctive content whereas in the original case they have consistent epistemic contents but contradictory subjunctive contents. In the InPierre case, Pierre's beliefs have both a contradictory subjunctive content *and* contradictory epistemic content. This is because one of the properties Pierre associates with Londres in InPierre is the property of being identical to the thing that plays the London role, and vice versa.

But both Lewis's proposal and Chalmers's proposal have a flaw; they both rely on the actual world supplying an object. There are cases which are, in all crucial respects, like the original Pierre case but in which the object is absent. We can construct cases in which we get the puzzling Pierre effect but in which the beliefs in question are empty.

Consider the case of Chloe. Chloe is at the pub chatting to her friend Jill. Jill tells Chloe about her suspicions that a witch has been terrorising the village and Chloe comes to believe that the witch that Jill was talking about lives alone. Jill leaves the pub and Chloe gets talking to Jeff. Jeff is also convinced a witch has been terrorising the village and Chloe comes to think that the witch that Jeff was talking about lives with an apprentice. By the end of the night, Chloe believes that one witch, the one Jill told her about, lives alone and another witch, the one Jeff told her about, lives with an apprentice. As it happens, Jeff and Jill had been talking earlier and Jill had given Jeff the idea that there is a witch terrorising the village in the first place (she had read about the witch in a newspaper

that neither Chloe nor Jeff have read). Jill and Jack both believe that there is only one witch terrorising the village. Witches do not exist.

If Chloe was told the causal history of the witch story back to Jill's telling Jeff about the witch, she would probably revise one of her beliefs thinking 'I see! The witch that Jill was talking about just is the witch that Jeff was talking about. I suppose she does not both live alone and live with an apprentice'. Chloe does not have this information about the proximate causal history of the story *in the same way* that Pierre in the original case does not have the information that London is Londres. Lewis (1986, 34) briefly mentions a similar case involving the story of Santa Claus.

There is a sense in which Chloe's beliefs concerning the witch are contradictory and a sense in which they are not. She has contradictory beliefs in the sense that she has beliefs that have a common focus that ascribe contradictory properties. She does not have contradictory beliefs concerning the witch in the sense that, just like Pierre, she is unaware that her beliefs have a common focus.

Consider a variation on the Chloe case that is just like the original Chloe case except that Jill's witch beliefs and Jeff's witch beliefs do not have a common causal history. They came to their beliefs independently. Call this case *LuChloe* (for lucky Chloe). Chloe's beliefs in the original case conflict in a way that they do not in *LuChloe* so there is a sense in which, in the original case, Chloe has conflicting beliefs about the witch.

Lewis's acquaintance-based approach struggles in the Chloe case. What should the defender say about the sense in which Chloe does have contradictory beliefs about the witch? She cannot answer that Chloe is acquainted with the witch in two different ways (via two different information bearing causal chains). The witch is missing from the actual world, so there are no information transmitting causal chains linking the object, if there is one, to Chloe. So the defender of a Lewis-style acquaintance-based view is unable to make the appropriate distinction between the original Chloe case and *LuChloe*. Therefore, Lewis's acquaintance-based approach leaves much to be desired when it comes to explaining the sense in which Pierre and Chloe contradict themselves and the sense in which they do not.

The defender of a Chalmers-style pluralism cannot account for the sense in which Chloe *does* have contradictory beliefs concerning the witch. Chloe's beliefs in the original case conflict in a way

that they do not in LuChloe so there is a sense in which, in the original case, Chloe has conflicting beliefs about the witch. The defender of a Chalmers-style pluralism captures the relevant distinction between the original Pierre case and LuPierre by claiming that in the original case Pierre's beliefs have contradictory subjunctive contents, whereas in LuPierre they have consistent subjunctive contents. The problem is that Chalmers cannot make the appropriate distinction between the original Chloe case and LuChloe in the same way. For Chalmers, beliefs that have no extension at the actual world do not have any non-trivial subjunctive content. In the original Chloe case and in LuChloe Chloe's beliefs are empty, there are no witches, so, by Chalmers's lights, her beliefs have no non-trivial subjunctive content. So, the Chalmers-style pluralist cannot account for the relevant difference between these cases by appealing to a difference in subjunctive content. A Chalmers-style pluralist approach can handle only a proper subset of the Pierre-type cases, the ones in which the relevant attitudes are not empty, but it struggles in Pierre-type cases in which the attitudes are empty. For this reason, Chalmers's approach to Kripke's puzzle leaves much to be desired.

The existence of cases like Chloe's case shows that any satisfactory account of Kripke's puzzle will be adaptable to cases in which the relevant attitudes are empty. We can get the puzzling Pierre effect even when there is no concrete object that the beliefs pick out at the actual world.

#### 7.4. My proposal

I propose that we ought to understand the puzzle in terms of intentional identity. That is to say, we should understand the original Pierre case in the following way: the sense in which Pierre has contradictory beliefs about London is captured by the fact that his beliefs are *in fact* g-related. The sense in which he does not have contradictory beliefs concerning London is captured by the fact that Pierre does not *believe* that they are g-related.

The existence of cases in which the object or objects to which the attitudes are directed are missing, provides a reason for thinking about things in this way. Theories of intentional identity are designed to handle cases in which the relevant attitudes are empty. Thus, understanding the puzzle in terms of g-relations promises to allow us to handle the Chloe case and Pierre case in a uniform way.

If this way of thinking about the case is correct, then we will need two things to provide an adequate

account of the puzzling cases. The first is a theory of the content of attitudes according to which it is possible for Pierre to believe that London is London (and that his attitudes about London are about London), without thereby believing that Londres is London (and that his attitudes about London are about Londres), even if London is Londres. The second is a theory of the g-relation, of when and why attitudes have a common focus. I will expand on these two requirements in turn.

#### 7.4.1. Beliefs about identity

We need to be able to distinguish the belief that Pierre does have, that London is London, from the belief that he apparently does not have, that London is Londres. I prefer the descriptivist theory of the content of attitudes, so I have a natural explanation of why and how this is possible. It is, however, crucial to note that my approach to Kripke's puzzle does not commit one to a descriptivist theory. A descriptivist theory allows us to make the appropriate distinctions between the relevant beliefs about identity, but so do other theories of mental content such as sophisticated Russellianism and Fregeanism.

Naïve Russellians notoriously have a hard time drawing the distinction between the belief that London is London and the belief that Londres is London. The problem is that when, in fact, London is Londres, the belief that London is London is equivalent to the belief that London is Londres. Both of these beliefs ascribe the same relation as standing between an object and itself. The naïve Russellian is thus forced to say that if Pierre believes that London is London, he also believes that Londres is London in the original case.

#### 7.4.2. A theory of intentional identity

As for a theory of intentional identity, I prefer the triangulation theory to alternatives, so below I will use triangulation theoretic tools to handle the puzzling cases. There are, however, other theories of intentional identity that could yield a solution to Kripke's puzzle.

For instance, one might adopt a transparent approach to intentional identity and apply it to solve Kripke's puzzle. If a transparent theory is right, then in both the original Pierre case and in the Chloe case, the agent's beliefs are both *about the same object* but they do not realise it. The transparent theorist can then distinguish the sense in which these agents have conflicting beliefs about

a city and a witch respectively in roughly the same way as I do, except that the g-relation would be understood in a different way.

A Chalmers-style pluralist picture fits well with a transparent theory of intentional identity. A defender of a Chalmers-style pluralism might, for example, posit the existence of mythical objects that exist at the actual world. I should emphasise that Chalmers does not take this option. This kind of Chalmers-style pluralist could then claim that for instance, Chloe's belief that Jill's witch lives alone does have non-trivial subjunctive content; its content is the set of worlds at which *that* mythical witch lives alone. Then Chloe's beliefs might have conflicting subjunctive content after all. This sort of move stands or falls with the transparent approach to intentional identity.

If I am right that Kripke's puzzle ought to be solved by engaging our favourite theory of intentional identity, then exactly how we *should* solve Kripke's puzzle comes down to which theory of intentional identity we should prefer. In chapters two through four inclusive I argued that the triangulation theory is the best theory of intentional identity on offer at the moment. I am, therefore, in a position to recommend my solution to Kripke's puzzle over similar solutions that employ other theories of intentional identity. That being said, even if the reader prefers another theory of intentional identity they can still take a lesson away from this chapter; we ought to understand Kripke's puzzle in terms of intentional identity.

### 7.5. My solution

My favoured solution to Kripke's puzzle involves combining a descriptivist theory of the content of attitudes with a triangulation theoretic understanding of the g-relation.

In the original Pierre case, Pierre lays out triangulation conditions for the putative target of his belief that London is pretty. He has beliefs about London's place in the world and what, in general, it would take for an attitude to be about London, from these beliefs we can extract some triangulation conditions that he lays out for London. The same is true, *mutatis mutandis*, for his Londres beliefs. According to the triangulation theory, two attitudes can be g-related, by an agent's standards, even if that agent does not believe that they are g-related (see section 7.8 of chapter five). In the original Pierre case, at least one triangulation condition that he lays out for the putative target of his belief that London is not pretty is, *in fact*, met by his belief that Londres is pretty. So these

two beliefs are g-related $_{\text{London}}^{\text{Pierre}}$ , they both focus on the same putative target, London. Presumably, if Pierre came to *believe* that these attitudes are g-related $_{\text{London}}^{\text{Pierre}}$  he would be disposed, if he had sensible views about cities, to revise at least one of his beliefs. In the original case, his beliefs are g-related (have a common focus) by his own standards, but he does not believe that they are. With this framework in place we can solve the puzzle.

Contrast this to Pierre's beliefs in the LuPierre case. Once again Pierre lays out some triangulation conditions for the putative target of his London belief. Ditto for his Londres belief. But in LuPierre, it looks as though his belief concerning Londres *does not* meet at least one of the relevant triangulation conditions. In LuPierre London and Londres are different cities. The relevant difference between Pierre's beliefs in the original case and his beliefs in LuPierre is whether his beliefs are g-related. In LuPierre they are not, but in the original case they are.

What, then, is the relevant difference between Pierre's beliefs in the original case and his belief in InPierre? The difference is that in InPierre, Pierre's beliefs are g-related *and* he believes that his attitudes are both about London. His Londres belief satisfies at least one of the triangulation conditions Pierre lays out for London, the putative target of his London belief so his London belief is g-related $_{\text{London}}^{\text{Pierre}}$  to his Londres belief. In the original case, Pierre does not believe that the two beliefs are g-related. So in this sense Pierre is not inconsistent in the same way as he is in InPierre.

What about the Chloe case? According to the triangulation theory, attitudes can have a common focus (be g-related) even when there is no object at that focus. So the triangulation theorist should handle the Chloe case in just the same way that they handled the Pierre case. Chloe lays out triangulation conditions for the putative targets of her witch beliefs. Her beliefs may well be, by her own standards, g-related even if she does not believe they are. We can then capture the relevant difference between the original Chloe case and LuChloe in a way that a defender of Chalmers's approach cannot. In the original case her beliefs are g-related, by her own standards, while in LuChloe they are not.

I have been talking as if the agent's own standards are crucial to the evaluation of these cases but this need not be the case. It may be that the conditions on intentional identity that are relevant are those of the reporter or the evaluator. Pierre's beliefs may still be g-related by the standards of the reporter or the evaluator and in spite of Pierre's not believing that they are.

We can handle similarly puzzling cases in a similar way. Consider, for instance, the case of Oedipus. Oedipus desires to marry Jocasta. As it happens, Jocasta is Oedipus' mother, but Oedipus does not realize this. In this case, does Oedipus desire to marry his mother? Again I think the answer should be yes and no. He desires to marry Jocasta and, by his own standards for sameness of target, his mother is the target of this desire. In this sense he does desire to marry his mother. On the other hand, he does not *believe* that Jocasta is his mother and, thus, he does not believe that his desire is about his mother. In this sense he does *not* desire to marry his mother.

Or suppose I am sitting in a restaurant and a man with a hood covering his face walks in. I believe that the hooded man has just walked into the restaurant. As it happens, the hooded man is my long lost brother. Do I believe that my long lost brother has just walked into the restaurant? Yes and no. By my own standards for sameness of target, my belief is about my long lost brother. This is the sense in which I do believe my brother has just walked into the restaurant. On the other hand, I do not *believe* that the hooded man is my brother and, thus, do not believe that my belief is about my brother even though I believe that it is about the hooded man.

Finally, suppose that Ralph is watching a spy at work in the shadows. He comes to believe that the spy he is watching is clever. As it happens, the spy is Bernard, though Ralph does not recognise him. In what sense is Ralph's belief that the spy he is watching is clever *about* Bernard? I answer that his belief satisfies at least one of the triangulation conditions he, Ralph, lays out for Bernard. So Ralph's belief is about Bernard, by his own standards, but he does not believe it is.

## 7.6. Conclusion

Does Pierre have contradictory beliefs concerning London? I answer yes and no. He does have contradictory beliefs in the sense that his beliefs are, in fact, g-related and ascribe conflicting properties. He does not have contradictory beliefs about London in the sense that he does not *believe* that his London belief and his Londres belief are g-related.

Unlike defenders of Lewis's approach or Chalmers's approach, those who adopt my approach can explain the puzzling phenomenon even in cases in which the target entity is missing. Since according to the triangulation theory attitudes can be g-related in spite of there being no concrete object at the actual world that they are both about, cases in which the target object is missing (like the



Chloe case) and cases in which it is not (like the original Pierre case) can be handled in the same way.

## CHAPTER 8: SUMMING UP AND DIRECTIONS FOR FUTURE RESEARCH

### 8.1. The first goal

In chapter one I stated the two main goals of this thesis. The first was to develop a good theory of *intentional identity*. In chapter four, after critiquing many of the main theories of intentional identity that are currently available in the literature, I presented what I take to be a strong contender for the best theory of intentional identity, the triangulation theory.

The triangulation theory has the advantage of delivering a uniform treatment of the relation of having a common focus across cases involving empty intentional attitudes and cases involving non-empty attitudes. This is an advantage it has over theories like those of Sainsbury (2010) and Geach, (1976, 314-317).

The triangulation theory is able to deliver correct verdicts about intentional identity in cases about which other theories yield incorrect verdicts. Unlike Sainsbury's theory and Perry's (2001, 14, 147-156) theory, the triangulation theory can deliver the correct verdicts in cases in which there is no causal link of the appropriate kind standing between the agents who have those attitudes. Unlike Dennett's (1968, 336-338, 341) theory, the triangulation theory can deliver the correct verdict that intentional identity is possible in spite of disagreement about the putative target. Unlike King's theory (1993, 65) and the lazy theory (1967, 630), the triangulation theory can deliver the correct verdict that intentional identity is possible even when the relevant agents have no beliefs about one another.

The triangulation theory is more complete than many of its rivals. For instance, it is more complete than Geach's (1976, 314-317) theory, Glick's (2012) theory, van Rooy's (2000, 170-178) theory and the transparent theories of Salmon (2005, 105–108), Parsons (1974, 577–578), and Saarinen (1978, 207–210).

The triangulation theory does not run into any of the problems I discussed in chapter two that face transparent theorists. What is more, my theory has the advantage of ontological and explanatory economy. It does not require us to posit any exotic objects or appeal to them in our explanation of

intentional identity.

Lastly, adopting a triangulation theoretic approach provides an elegant and comprehensive solution to the puzzle that arises out of the cases presented by Edelberg (1986, 1992). The theories of Geach, Salmon, Parsons, and Sainsbury struggle to handle Edelberg-style cases, see section 3.13. Glick (2012, 395-396) provides a partial solution to Edelberg's puzzle but, as I argued in section 3.14, Glick's strategy is importantly limited in a way that the triangulation theoretic solution is not, see section 4.6.

Intentional identity is a puzzling phenomenon and it has not been discussed as much as it deserves to be. I have taken a reasonably good stab at accounting for it.

## 8.2. The second goal

The second goal of this thesis was to use the tools developed in the service of finding a good theory of intentional identity to shed light on other debates concerning intentional attitudes. Triangulation theoretic tools I developed in chapter four to shed light on other debates concerning intentional attitudes and suggest a new approach to intentional attitudes.

The descriptivist theory of attitude content has many attractive features. It provides an elegant, coherent and natural account of many philosophically puzzling phenomena, empty attitudes, beliefs about identity, and so on. Even Kripke (1980, 5), though he abandons much of the 'complex of ideas, emanating from Frege and Russell', praises them for their naturalness, power, and 'marvelous internal coherence'.

But the descriptivist theory of the content of attitudes also faces serious and pressing challenges. There are concerns about how fragile attitudes are on a descriptivist story. The fact that agents often believe, desire or fear the same thing seems to put pressure on the descriptivist theory of content, see section 5.5. It also seems as if it is hard to make sense of phenomena such as communication, agreement, and disagreement on a descriptivist account of content. Cousins of arguments against the description theory of names are also often leveled against the descriptivist account of attitude content.

In chapter six I argued that many of the most serious objections can be answered if the descriptivist helps herself to triangulation theoretic tools. The descriptivist approach to the content of attitudes

is strengthened by being combined with a triangulation theoretic approach to subject matter.

As we saw in chapter seven, the triangulation theory also helped us reach a novel solution to Kripke's puzzle. The triangulation theoretic tools shed light *directly* on that corner of the literature.

The triangulation theory also suggests a systematic approach to questions of intentionality and mental content. There is a thought that appears to sit in the background of many debates about mental content but which is not often explicitly stated. The thought is that explaining certain phenomena *requires* appealing to external constraints on the content of attitudes. 'Some form of externalism has to be true' an externalist might say 'otherwise we can't account for agents believing, fearing, desiring the same thing, genuine disagreement, communication, etc.' In the second part of this thesis I have taken some steps toward dispelling this line of thought. The triangulation theory allows us to explain these phenomena in a way that does not require external determinants of the content of attitudes.

This suggests a general systematic approach to intentionality that has been partly implemented in this thesis. Neither the descriptivist theory of the content of attitudes nor the triangulation theory require external constraints on mental content. I would like to suggest a broadly internalist/individualistic approach to intentionality. Since the 1980s or so, externalism about mental content (roughly, the doctrine that what an agent believes, fears, desires, etc. does not supervene on what is going on inside the holder of the attitude's skin) has been almost the orthodox view among philosophers of mind and language. Externalism is somewhat costly, as I argued in chapter five, but many are willing to pay that cost because they think that externalism is the only way to explain certain phenomena such as communication, agreement and disagreement. By showing that these phenomena can be accounted for *without* appealing to external constraints on the content of intentional attitudes, the discussion in the second part of this thesis has undermined this kind of motivation for externalism. This is a victory for those of us who reject the usual brands of externalism, even if it is not a decisive victory.

By combining the descriptivist theory of content with a triangulation theoretic account of subject matter, we have made some, albeit limited, progress toward a unified approach to questions of mental content. The moves made in the second part of the thesis are instances of looking for guidance on matters of mental content and intentionality in the details of an agent's narrowly construed

psychological state. We look for conditions on triangulation, not out in the world, but in the heads of agents. Just as the content of attitudes, on the descriptivist story I defend, ought to be extracted from the details of the holder of the attitude's narrowly construed psychological state. Jackson (2003, 2004, 2010) is one philosopher who, in his relatively recent work, has defended a broadly individualistic approach to mental content. The second part of this thesis does something to show the potential of this kind of approach to mental content.

### 8.3. Other applications

What are some of the implications the material in this thesis has for other debates? One obvious application of the ideas presented in this thesis is to the philosophy of language. The heart of the triangulation theory is agents' representations *of* the target and their views about what it would take for an attitude (a representational state) to be about that target. Triangulation conditions are extracted from an agent's beliefs about what it would take for other attitudes, which we ought to understand as representational, to represent a given target. It is natural to think that agents would also have beliefs about what it would take for a term in a language to represent the target entity. We might be able to develop a theory of the linguistic content and of the subject matter of linguistic utterances that is somewhat similar to the theory of mental content and the subject matter of attitudes presented in this thesis.

As I mentioned in chapter one, the ideas in this thesis have implications for more general questions of subject matter. If the triangulation theory is right, we have a way of systematically making sense of and evaluating claims that this or that argumentative move *changes the subject*. What is more, the triangulation theory highlights the utility of making explicit the beliefs about what it would take for the subject matter of a discourse to change. Most of the time these conditions are not discussed explicitly. By revealing their importance, the triangulation theory will hopefully push people to make explicit the beliefs about conditions on subject matter individuation that underpin their judgements about when the subject of discourse changes.

Another interesting line of future research would involve mapping the material in this thesis to the debate concerning different kinds of disagreement, such as verbal disagreement, non-verbal disagreement and so on. I have in mind a picture of disagreement that is closely tied to the notion of

subject matter. Sometimes the subject matter of a dispute is the semantics of words, whereas sometimes the subject matter is the thing those words are supposed to pick out. Sometimes the dispute concerns how the terms *should* be used or about how the target of the discourse *should* be conceived. The triangulation theory will help us systematically distinguish these different forms of disagreement by providing a powerful and flexible way of understanding questions of subject matter individuation.

Another application of the triangulation theory is to the semantics of transitive verbs. Consider, for instance, (16).

(16) Jack and Jill both sought the fountain of youth.

There is a puzzle about how Jack and Jill can both seek something that does not exist so as to make (16) come out true even in cases in which there is no fountain of youth. There is also a question regarding the relationship between how Jack and Jill take the thing they are seeking to be and the applicability of transitive verb-based sentences to their attitudes. Triangulation theoretic tools can help here as well. Roughly speaking, seeking the same thing might be understood in terms of directing one's actions at a putative target. The way the agents take the putative target to be will yield some triangulation conditions. We can then come up with a schema analogous to T3 that will concern directed actions (seeking, avoiding, etc.) that will, in turn, deliver answers about when instances of seeking and avoiding have a common target. This is only a rough gloss but hopefully it is clear how triangulation theoretic tools could shed light on debates concerning transitive verbs.

Some debates in the history and philosophy of science concerning the individuation of subject matter can also be illuminated by triangulation theoretic tools. For instance, there is a debate about when scientific theories have the same subject matter across time. Did Einstein and Newton have beliefs that had a common focus when they theorized about space? How can the subject matter of scientific theories stay stable across time in spite of significant changes in how the target phenomenon is taken to be? The triangulation theory can shed light on these questions by supplying a way of understanding conditions on the individuation of subject matter.

Triangulation theoretic tools can also shed light on some debates in meta-ethics. When two people who disagree about what 'the good' amounts to have beliefs about 'the good' do their beliefs have a common focus? One complaint that is made against certain meta-ethical theories is that they

cannot make sense of genuine ethical disagreement. This is because, according to that meta-ethical theory, parties to the dispute do not mean the same thing when they use terms that are crucial to the debate such as 'good' and 'right'. In future work I plan to use triangulation theoretic tools to defend a certain kind of meta-ethical subjectivism against a charge of this sort. Triangulation theoretic resources allow this kind of worry to be dispelled. It may be that though the terms that the parties to the debate are using mean different things in their respective mouths, they have a common focus nonetheless. In this way agents meaning different things by the term 'good', for instance, is not required for them to have a genuine disagreement concerning the good.

## REFERENCES

- Braddon-Mitchell, David, and Frank Jackson. 1996. *The philosophy of mind and cognition*. Oxford, OX, UK ; Cambridge, Mass., USA: Blackwell Publishers.
- Braun, David. 1998. "Understanding belief reports." *Philosophical Review* 107 (4):555-595.
- Braun, David. 2001a. "Russellianism and Explanation." *Noûs* 35 (s15):253-289.
- Braun, David. 2001b. "Russellianism and prediction." *Philosophical Studies* 105 (1):59 - 105.
- Braun, David. 2012. "Hob, Nob, and Mythical Witches." In *Reference and Referring: Topics in Contemporary Philosophy*, edited by Joseph Keim Campbell Michael O'Rourke and Harry S Silverstein. Cambridge, MA: MIT Press.
- Burge, Tyler. 1979. "Individualism and the mental." *Midwest Studies in Philosophy* 4 (1):73-122.
- Burge, Tyler. 1986. "Individualism and psychology." *Philosophical Review* 95 (January):3-45.
- Chalmers, David J. 2002a. "The components of content." In *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers. Oxford University Press.
- Chalmers, David J. 2002b. "On sense and intension." *Philosophical Perspectives* 16 (s16):135-82.
- Church, Alonzo. 1951. *A Formulation of the Logic of Sense and Denotation*. Edited by P H. Kallen, Henle and S, Langer, *Structure, Method, and Meaning: Essays in Honor of Henry M. Sheffer*. New Yord: Liberal Arts Press.
- Cumming, Samuel. 2014. "Indefinites and intentional identity." *Philosophical Studies* 168 (2):371-395.
- Dennett, Daniel. C. 1968. "Geach on intentional identity." *Journal of Philosophy* 65 (11):335-341.
- Devitt, Michael. 1996. *Coming to our senses : a naturalistic program for semantic localism, Cambridge studies in philosophy*. Cambridge ; New York: Cambridge University Press.
- Donnellan, Keith S. 1970. "Proper Names and Identifying Descriptions." *Synthese* 21 (3/4):335-358.
- Dummett, Michael. 1973. *Frege: Philosophy of Language*: Duckworth.
- Edelberg, Walter. 1986. "A new puzzle about intentional identity." *Journal of Philosophical Logic* 15 (1):1 - 25.



- Edelberg, Walter. 1992. "Intentional identity and the attitudes." *Linguistics and Philosophy* 15 (6):561 - 596.
- Edelberg, Walter. 1995. "A Perspectivalist Semantics for the Attitudes." *Noûs* 29 (3):316-342. doi: 10.2307/2215602.
- Fodor, Jerry A., and Ernest LePore. 1992. *Holism: A Shopper's Guide*. Vol. 43: Blackwell.
- Fodor, Jerry A., and Ernest LePore. 2002. *The Compositionality Papers*: Oxford University Press.
- Frege, Gottlob. 1892. "Über Sinn und Bedeutung." *Zeitschrift für Philosophie Und Philosophische Kritik* 100 (1):25--50.
- Friend, Stacie. 2014. "Notions of Nothing." In *Empty Representations: Reference and Non-Existence*, edited by Manuel Garca-Carpintero and Genoveva Mart. Oxford University Press.
- Geach, Peter T. 1967. "Intentional identity." *Journal of Philosophy* 64 (20):627-632.
- Geach, Peter T. 1976. "Two kinds of intentionality?" *The Monist* 59 (July):306-320.
- Glick, Ephraim. 2012. "A Modal Approach to Intentional Identity." *Noûs* 46 (3):386-399.
- Heim, Irene. 1982. *The Semantics of Definite and Indefinite Noun Phrases*.
- Heim, Irene. 1990. "E-type pronouns and donkey anaphora." *Linguistics and Philosophy* 13 (2):137--77.
- Jackson, Frank. 1998. "Reference and Description Revisited." *Noûs* 32 (12):201-218.
- Jackson, Frank. 2003. "Narrow content and representation--or twin earth revisited." *Proceedings and Addresses of the American Philosophical Association* 77 (2):55-70.
- Jackson, Frank. 2004. "Why We Need A-Intensions." *Philosophical Studies* 118 (1/2):257-277.
- Jackson, Frank. 2010. *Language, names, and information, The Blackwell/Brown lectures in philosophy*. Malden, MA: Wiley-Blackwell.
- Joyce, Richard. 2008. "Morality, schmorality." In *Morality and Self-Interest*, edited by Paul Bloomfield. Oxford University Press.
- King, Jeffrey C. 1993. "Intentional identity generalized." *Journal of Philosophical Logic* 22 (1):61 - 93.
- Kripke, Saul. 1977. "Speaker's Reference and Semantic Reference." *Midwest Studies In Philosophy* 2 (1):255-276. doi: 10.1111/j.1475-4975.1977.tb00045.x.

- Kripke, Saul A. 1979. "A puzzle about belief." In *Meaning and Use*, edited by A. Margalit, 239--83. Reidel.
- Kripke, Saul A. 1980. *Naming and necessity*. Cambridge, Mass.: Harvard University Press.
- Kripke, Saul A. 2013. *Reference and Existence. The John Locke Lectures*: Oxford University Press.
- Kroon, Frederick. 1987. "Causal descriptivism." *Australasian Journal of Philosophy* 65 (1):1-17.
- Kroon, Frederick. 2009. "Names, Plans, and Descriptions." In *Conceptual analysis and philosophical naturalism*, 214-242. Cambridge, Mass.: MIT Press.
- Lanier, William. 2014. "Intentional identity and descriptions." *Philosophical Studies* 170 (2):289-302.
- Lewis, David. 1974. "Radical interpretation." *Synthese* 27 (July-August):331-344.
- Lewis, David. 1981. "What puzzling Pierre does not believe." *Australasian Journal of Philosophy* 59 (3):283 – 289.
- Lewis, David. 1983. "New work for a theory of universals." *Australasian Journal of Philosophy* 61 (December):343-377.
- Lewis, David. 1984. "Putnam's paradox." *Australasian Journal of Philosophy* 62 (3):221-236. doi: 10.1080/00048408412340013.
- Lewis, David. 1986. *On the Plurality of Worlds*: Blackwell Publishers.
- Lewis, David. 1993. "Many, but almost one." In *Ontology, Causality, and Mind: Essays on the Philosophy of D. M. Armstrong*, edited by Keith Cambell, John Bacon and Lloyd Reinhardt, 23-38. Cambridge University Press.
- Lewis, David. 1997. "Naming the colours." *Australasian Journal of Philosophy* 75 (3):325-42.
- Loar, Brian. 1976. "The semantics of singular terms." *Philosophical Studies* 30 (6):353 - 377.
- Manning, Luke. 2015. "No Identity Without an Entity." *Pacific Philosophical Quarterly* 96 (2):279-305. doi: 10.1111/papq.12074.
- Mates, Benson. 1952. "Synonymity." In *Semantics and the Philosophy of Language*, edited by Leonard Linsky, 286. Urbana: University of Illinois Press.
- McKay, Thomas. 1981. "On proper names in belief ascriptions." *Philosophical Studies* 39 (3):287-303.
- Mckinsey, Michael. 1986. "Mental anaphora." *Synthese* 66 (1):159 - 175.

- Nolan, Daniel. 1996. "Recombination unbound." *Philosophical Studies* 84 (2-3):239 - 262.
- Nolan, Daniel, and Alexander Sandgren. 2014. "Creationism and cardinality." *Analysis* 74 (4):615-622.
- Parfit, Derek. 2011. *On What Matters*: Oxford University Press.
- Parsons, Terence. 1974. "A Prolegomenon to Meinongian Semantics." *The Journal of Philosophy* 71 (16):561-580.
- Perry, John. 1979. "The Problem of the Essential Indexical." *Noûs* 13 (1):3-21. doi: 10.2307/2214792.
- Perry, John. 2001. *Reference and Reflexivity*: Stanford: CSLI Publications.
- Putnam, Hilary. 1975. "The meaning of 'meaning'." *Minnesota Studies in the Philosophy of Science* 7:131-193.
- Quine, Willard van Orman. 1948. "On What There Is." *Review of Metaphysics* 2 (5):21--36.
- Reddam, Paul. 1982. *Pragmatics and the Language of Belief*: Dissertation.
- Russell, Bertrand. 1905. "On Denoting." *Mind* 14 (56):479-493.
- Russell, Bertrand. 1910-1911. "Knowledge by Acquaintance and Knowledge by Description." *Proceedings of the Aristotelian Society* 11:108-128.
- Saarinen, Esa. 1978. "Intentional Identity Interpreted: A Case Study of the Relations among Quantifiers, Pronouns, and Propositional Attitudes." *Linguistics and Philosophy* 2 (2):151-223. doi: 10.2307/25000983.
- Sainsbury, Mark. 2005. *Reference Without Referents*: Clarendon Press.
- Sainsbury, Mark. 2010. "Intentionality Without Exotica." In *New Essays on Singular Thought*, edited by Robin Jeshion. Oup Oxford.
- Sainsbury, Mark, and Michael Tye. 2011. "I—An Originalist Theory of Concepts." *Aristotelian Society Supplementary Volume* 85 (1):101-124.
- Sainsbury, Mark, and Michael Tye. 2012. *Seven Puzzles of Thought: And How to Solve Them: An Originalist Theory of Concepts*: Oup Oxford.
- Salmon, Nathan. 1986. *Frege's Puzzle*: Ridgeview.
- Salmon, Nathan. 1989. "Illogical belief." *Philosophical Perspectives* 3:243-285.
- Salmon, Nathan. 1990. "A millian heir rejects the wages of sinn." In *Propositional Attitudes: The Role of Content in Logic, Language, and Mind*, edited by C. A. Anderson and J. Owens, 215-247. CslI Publications.

- Salmon, Nathan. 1998. "Nonexistence." *Noûs* 32 (3):277-319.
- Salmon, Nathan. 2005. "Mythical objects." In *Metaphysics, Mathematics, and Meaning*, 91-108. New York: Oxford University Press. Original edition, 2002.
- Salmon, Nathan. 2006. "The resilience of illogical belief." *Noûs* 40 (2):369–375.
- Saul, Jennifer. 1997. "Substitution and simple sentences." *Analysis* 57 (2):102–108.
- Saul, Jennifer. 1999. "The Best of Intentions: Ignorance, Idiosyncrasy, and Belief Reporting." *Canadian Journal of Philosophy* 29 (1):29 - 47.
- Schroeter, Laura. 2008. "Why be an anti-individualist?" *Philosophy and Phenomenological Research* 77 (1):105-141.
- Sider, Theodore. 2009. "Ontological realism." In *Metametaphysics: New Essays on the Foundations of Ontology*, edited by David John Chalmers, David Manley and Ryan Wasserman. Oxford University Press.
- Slater, Hartley. 2000. "The grammar of the attitudes." In *Reference and Anaphoric Relations*, edited by Klaus von Heusinger and Urs Egli, 183--190. Kluwer.
- Soames, Scott. 1987. "Direct reference, propositional attitudes, and semantic content." *Philosophical Topics* 15 (1):47-87.
- Sorensen, Roy A. 1988. *Blindspots*. Vol. 25: Oxford University Press.
- Stalnaker, Robert. 1978. "Assertion." *Syntax and Semantics (New York Academic Press)* 9:315-332.
- Stalnaker, Robert. 1984. *Inquiry*: Cambridge University Press.
- Stalnaker, Robert. 1988. "Belief attribution and context." In *Contents of Thought*, edited by Robert H. Grimm and D. D. Merrill, 140--156. University of Arizona Press.
- Stalnaker, Robert. 2003. *Ways a World Might Be: Metaphysical and Anti-Metaphysical Essays*: Oxford University Press.
- Thomasson, Amie L. 1996. "Fiction and intentionality." *Philosophy and Phenomenological Research* 56 (2):277-298.
- Thomasson, Amie L. 1999. *Fiction and Metaphysics*: Cambridge University Press.
- Tye, Michael. 1978. "The puzzle of Hesperus and Phosphorus." *Australasian Journal of Philosophy* 56 (3):219 – 224.

- van Rooy, Robert. 2000. "Anaphoric Relations Across Attitude Contexts." In *Reference and Anaphoric Relations*, edited by Klaus von Heusinger and Urs Egli, 157-181. Springer Netherlands.
- van Rooy, Robert, and Thomas Ede Zimmermann. 1996. "An externalist account of intentional identity." Proceedings in the Konstanz Workshop "Reference and Anaphorical Relations", Universität Konstanz.
- Weatherson, Brian. 2003. "What good are counterexamples?" *Philosophical Studies* 115 (1):1-31.
- Weber, Clas. 2013. "Centered communication." *Philosophical Studies* 166 (1):205-223.
- Zimmermann, Thomas Ede. 1999. "Remarks on the Epistemic Role of Discoarse Referents." In *Logic, Language and Computation, Volume 2*, edited by Lawrence S. Moss, Jonathan Ginzburg and Maarten de Rijke. Center for the Study of Language and Inf.
- Zvolenszky, Zsófia. 2014. "Artifactualism and Authorial Creation." *Proceedings of the European Society for Aesthetics*.
- Zvolenszky, Zsófia. 2012. "Against Sainsbury's Irrealism About Fictional Characters: Harry Potter as an Abstract Artifact." *Hungarian Philosophical Review (Magyar Filozófiai Szemle)* (4):83-109.