

Imagination, Content, and the Self

by

Dilip Ninan

BA (Hons), University of Western Ontario (1999)

BPhil, Oxford University (2001)

MSc, Oxford University (2002)

Submitted to the Department of Linguistics and Philosophy
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2008

© Massachusetts Institute of Technology 2008. All rights reserved.

Author
Department of Linguistics and Philosophy
May 21, 2008

Certified by
Robert C. Stalnaker
Laurence S. Rockefeller Professor of Philosophy
Thesis Supervisor

Accepted by
Alex Byrne
Professor of Philosophy
Chairman, Department Committee on Graduate Students

Imagination, Content, and the Self

by

Dilip Ninan

Submitted to the Department of Linguistics and Philosophy
on May 21, 2008, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

I begin with a discussion of two ways of imagining something: ‘from the inside’ and ‘from the outside’. My interest in this topic is two-fold: First, I want to see what studying this topic can teach us about the nature of mental content—in particular, about the content of *de se* and *de re* thoughts. Second, I want to formulate an account of this distinction which will help us understand the role these two types of imagining play in philosophical thought experiments about personal identity over time. The first three chapters of this dissertation focus on the first set of issues, issues of imagination and content. Chapter 4 extends and applies some of these insights to a puzzle about personal identity over time.

Thesis Supervisor: Robert C. Stalnaker
Title: Laurance S. Rockefeller Professor of Philosophy

Acknowledgments

There are many people – call them ‘enablers’ – who helped me to write this dissertation.

Thanks to the MIT philosophy faculty, graduate students, and staff for creating the Platonic Form of a philosophy department. Bob Stalnaker was a supportive and thoughtful supervisor, and his feedback substantially improved many of the key ideas developed here. The other members of my committee, Alex Byrne, Richard Holton, and Steve Yablo, all provided excellent feedback on various parts of the dissertation, as well as much-needed encouragement. I also benefitted from interacting with the linguistics part of the Department of Linguistics and Philosophy, mostly in the person of Kai von Fintel. I arrived at some of the main ideas in this dissertation after reflecting on things I learned in Kai’s LSA Summer Institute class on pragmatics.

I was lucky to have Seth Yalcin as a friend and classmate during my time at MIT. I learned an enormous amount from our conversations on a wide range of issues, philosophical and linguistic. I thank him for that, and for help with various parts of the dissertation. Sarah Moss read more of this dissertation than anyone not professionally obligated to do so, and gave me excellent comments on virtually all of the ideas presented here. A number of sessions on the whiteboard with Alejandro Pérez Carballo helped me a great deal in thinking through earlier versions of Chapters 2 and 3, and several conversations with Caspar Hare helped me to get clearer on the material presented in Chapter 4.

Different portions of this dissertation have benefitted from comments from Mahrad Almotahari, Lauren Ashwell, Rachael Briggs, David Etlin, Ephraim Glick, Leah Henderson, Vann McGee, Agustín Rayo, Paolo Santorio, Eric Swanson, Judith Jarvis Thomson, Valentina Urbanek, and Kenny Walden. Some of this material was presented to the 2006 Brown, Oxford, and Arché philosophy graduate conferences, the 2007 Meeting of the Canadian Philosophical Association, the M.A.T.T.I. Group at MIT, the Harvard-MIT Friends and Eminees Group, and the MIT Work-in-Progress Seminar. Thanks to those audiences, and especially to Antony Eagle, Danny Goldstick, Jonathan Ichikawa, Ted Sider, Bruno Whittle, and Timothy Williamson.

For beds, futons, sofas, meals, and friendship in Cambridge this last academic year, thanks to Lauren Ashwell, Alejandro Pérez Carballo and Katia Vavova, Marco Gualtieri, Rae Langton and Richard Holton, and

Chris and Ulcca Joshi Hansen.

Special thanks are due to my parents, Abraham and Leela Ninan, for helping in ways far too numerous to mention.

Finally, thanks to Gillian Sinnott, closest companion, for everything.

Contents

1 Imagination and the <i>De Se</i>	9
1.1 Inside-outside as centered-uncentered	10
1.2 A linguistic argument	13
1.3 The problem of the ‘impersonal- <i>de se</i> ’	22
1.4 Towards a solution	26
2 Imagination, Descriptivism, and the <i>De Re</i>	31
2.1 The Naive Theory	33
2.2 Lewis on <i>de re</i> attitudes	34
2.3 The problem of counterfactual attitudes	37
2.4 A two-dimensional account	41
2.5 <i>De se</i> imagining again	45
2.6 Attitude ascription	50
3 Imagination, Sequenced Worlds, and the <i>De Re</i>	57
3.1 Identification and stipulation	58
3.2 Identity puzzles and acquaintance relations	65
3.3 Generalizing the theory	70
3.4 Imagining and sequenced worlds	71
3.5 Attitude ascription again	73
4 Imagination, Possibility, and Personal Identity	77
4.1 Introduction	77
4.2 Supervenience	79
4.3 The puzzle	81
4.4 Imagination and possibility	89
4.5 The solution	96
4.6 Something like the Simple View	104
A Semantics for centered possibility	107

Chapter 1

Imagination and the *De Se*

Consider the difference between imagining skiing down a steep hill and imagining *that* you are skiing down the same hill. In the first case, you imagine what it would be like to actually be there, on the hill. The imaginative scenario is viewed through your imaginary ski goggles, and includes a view of the tips of your skis and the white snow before yourself; it may be accompanied by imagining being cold and full of adrenaline. In the other case, you take a god's-eye view on the scenario and 'see' yourself zooming down the hill, as if you were watching a film of yourself skiing.

The first type of imagining is 'from the inside' or from the first-person point of view. The second type of imagining is 'from the outside' or from the third-person point of view.¹ In the above paragraph, we used two heuristics for explaining the distinction: a linguistic characterization, between imagining *doing* or *being* or *experiencing* something versus imagining *that* something is the case; and a distinction between the characteristic visual imagery associated with each type of imagining.

Thinking about visual imagery makes the distinction vivid, but it is potentially misleading. The distinction is not merely a distinction between two kinds of perceptual imagery. I can imagine from the inside being Napoleon, sitting in a pitch-black and completely quiet room, and I can do this by visualizing a featureless black expanse; I can also imagine that the universe contains nothing by visualizing a featureless black expanse. The former imagining is from the inside, the latter is not. I can also imagine a scenario from a particular visual point of view, without imagining being

¹The inside-outside distinction (or something like it) appears, in different guises, in many places in the literature. See, e.g., Williams (1966), Nagel (1974, n.11), Peacocke (1985), Walton (1990, 28-35), Shoemaker (1994), Velleman (1996), Hill (1997), Martin (2002), Higginbotham (2003), and Recanati (2007a,b).

the individual *at* that point of view. For example, I can imagine the visual experience of someone skiing down a hill, without imagining actually being the one doing the skiing, controlling the movements of the skier's body.²

An imagining from the inside may exploit both types of imagery. For example, I may imagine being a race-car driver in the following way (Williams, 1966, 38): I imagine feeling tense, my body jolted around, hands clasped tightly around the steering wheel; then I picture a car – my car – zooming across the finish line; and then my garlanded neck and beaming face on the front of *F1 Racing* (world's best-selling grand prix magazine). It is natural to describe my entire imaginative episode as “imaging from the inside being a race-car driver” even though not *all* of the associated imagery is from the driver's visual point of view.

There is a larger question looming the in background: what is the relationship between imagery and imagination? We won't have much to say about this, but the two previous paragraphs indicate some of the subtleties involved. We can, I think, begin the search for a more precise characterization of the inside-outside distinction, even if a number of questions about its precise nature are left open for the moment.

1.1 Inside-outside as centered-uncentered

Focus on imagining from the outside. Pretend for a moment that I have super-powers of imagining, and that I can imagine a scenario in *maximally specific* detail—that is, suppose I can imagine a whole world, fixing every last detail of the imagined world. Suppose I imagine from the outside a situation in which the New England Patriots win the Super Bowl over the New York Giants, and I imaginatively fix every possible detail of this situation. How could we characterize the *content* of this imagining?

The content of my imagining specifies in complete detail a certain possible situation, a way things could be. It is natural, then, to represent the content of my imagining by a certain possible world: the possible world which is exactly as my imagining specifies the world to be (there is only one, since my imagining is maximally specific). Call that world '*w*': *w* represents the content of my imagining from the outside that the Patriots win the Super Bowl.

²Thanks to Steve Yablo for pointing this out to me. Sometimes I will talk about ‘imagining such-and-such from *x*'s point of view’. But this should not be taken simply to mean ‘imagining things from *x*'s visual perspective’; it should be taken to mean ‘imagining being *x*'. I think this is acceptable usage.

1. Imagination and the *De Se*

Continue to pretend I have super-powers of imagining. Suppose now I undertake a different imaginative task: I imagine *exactly* the same scenario described above – the very same game, the very same world – but I imagine that scenario from the point of view of one of its participants. I imagine being Tom Brady, the quarterback of the New England Patriots, in the scenario described above. I imagine completing touchdown passes, leaving the field with victory in hand, falling into the arms of my super-model girlfriend.

How could we represent the content of *this* imagining? Here's an idea: represent the content as the *pair* consisting of the possible world w and Tom Brady, i.e. $\langle w, \text{Tom Brady} \rangle$. That w is the first member of the pair represents the fact that the *objective scenario* I'm imagining corresponds to the possible world w , and so is the very same objective scenario I earlier imagined from the outside. That Tom Brady is the second member of the pair represents the fact that I am looking at w from Brady's point of view: I am imagining being Tom Brady in that objective scenario.

Note that I could imagine the very same scenario from another point of view: I could imagine being Eli Manning, the quarterback for the New York Giants. I imagine being sacked by the Patriots' front line, losing the game, and then being pilloried in the *New York Post*. The content of this imagining could be represented as $\langle w, \text{Eli Manning} \rangle$, representing the fact that I am again imagining the very same objective scenario (w) as I did in the two imaginings described above, but this time I am viewing things from Manning's point of view.

So this account enables us to represent the fact that in all three cases I imagine the same objective scenario: in the first case, the content of the imagining is simply the possible world w , reflecting the fact that I'm imagining that scenario from the outside and so from no one's point of view; in the second case, the content of my imagining is $\langle w, \text{Tom Brady} \rangle$ representing the fact that I'm viewing the scenario from Brady's point of view; in the third case, the content is $\langle w, \text{Eli Manning} \rangle$, representing the fact that I'm viewing the scenario from Manning's point of view.

A note on terminology: a pair consisting of a possible world w and an inhabitant x of that world is called a *centered world*: x is said to be *the center*, and $\langle w, x \rangle$ is *centered on* x .³

Of course, I do not have super-powers of imagining and cannot imagine

³Often, a centered world is taken to be a world-time-individual triple. I will for the most part ignore this extra complexity since I will not be talking about the temporal properties of attitudes. The centered worlds terminology is due to Quine (1969).

things in maximally specific detail. Instead of representing the content of an imagining from the outside with a single possible world, we can represent it by a *set* of possible worlds. When I imagine from the outside that the Patriots win the Super Bowl, the content of my imagining is the set of possible worlds in which the Patriots win the the Super Bowl. My imagining is unspecific on ever so many details. For example, I do not imaginatively ‘settle’ whether the game is played on a Friday or on a Sunday. If p is the set of possible worlds compatible with what I imagine, then in some possible world w in p , the game is played on a Friday, and in some possible world w' in p , it is played on a Sunday. Call a set of possible worlds an *uncentered content*, or an *uncentered proposition*, or a *possible worlds proposition*.

Similarly, we can represent the content of an imagining from the inside as a set of centered worlds. I may imagine being one of the fans, watching the game from the stands. My imagining doesn’t settle whether I am imagining being forty years old or fifty years old. If p is the set of centered worlds compatible with what I imagine, then some of the centered worlds $\langle w, x \rangle$ in p are such that x is forty years old in w , others $\langle w', x' \rangle$ are such that x' is fifty years old in w' . Call a set of centered worlds a *centered content*, or a *centered proposition*.

Note that often when we talk about *the* content of an imagining, we are speaking somewhat loosely. If I imagine from the outside that the Patriots win the Super Bowl, my imagining will likely be somewhat richer than what is captured by the set of possible worlds in which the Patriots win the Super Bowl. My imagining may be somewhat committal on, for example, what the weather is like, whether the players are wearing helmets, whether there are more than one hundred people at the game, and so on. Often when we say that *the* content of my imagining is the proposition p , we merely mean that the strongest proposition I imagine entails (is included in) p . Strictly speaking, we should identify *the* content of my imagining with the strongest proposition I imagine, i.e. the set of possible or centered worlds compatible with what I imagine. But where no harm will be done by speaking loosely, I will continue to do so.

We can state our account of the inside-outside imagination distinction as follows:

Uncentered Imagination

The content of an imagining from the outside is a set of possible worlds.

For all imaginings I , I is an imagining from the outside iff I 's content is a set of possible worlds.

Centered Imagination

The content of an imagining from the inside is a set of centered worlds.

For all imaginings I , I is an imagining from the inside iff I 's content is a set of centered worlds.

There are, of course, well known problems for coarse-grained views of content, i.e views which take mental contents to be sets of possible situations. The following are both consequences of our account:

- Whenever I imagine anything, I imagine everything that is necessarily true.
- Whenever I imagine p , I imagine everything that follows from p .

These are problems for both **Uncentered Imagination** and **Centered Imagination**. These problems, in their full generality, will not be the focus of this dissertation. However, we will be looking at certain more specific versions of these problems, and we will refine our account accordingly. These refinements which will help solve at least some of the problems that face coarse-grained accounts of content. These matters will be discussed in due course (see Chapters 2 and 3).⁴

1.2 A linguistic argument

It is not surprising that a natural and intuitive account of the inside-outside distinction identifies it with the centered-uncentered distinction. Lewis (1979) uses centered worlds to give an account of the distinction

⁴In a sense, in moving from possible worlds to centered worlds, we are already beginning to chip away at these problems. For example, the set of possible worlds in which I am hungry is the set of possible worlds in which Dilip is hungry. So, on the possible worlds theory, if I believe that I am hungry, I must believe that Dilip is hungry (and vice versa), which seems wrong: if I'm a hungry amnesiac, I might believe the former without believing the latter. This problem seems to be an instance of the second problem mentioned above. Taking contents to be sets of centered worlds helps solve this particular instance of that problem, because the set of centered worlds in which the center is hungry is distinct from (and doesn't entail) the set of centered worlds in which Dilip is hungry. So I can believe one without believing the other.

between *de se* and *de dicto* attitudes. Since the inside-outside distinction is a first-person/third-person distinction, it is natural to think it has some connection to the *de se-de dicto* distinction. In this section, I want to show that we can give a more powerful argument for our account of the inside-outside distinction by exploiting this connection. I will introduce Lewis's account of the *de se-de dicto* distinction, motivate a semantic theory of attitude ascription based on Lewis's account, and then show how that semantic theory yields an argument that supports Centered and Uncentered Imagination.

1.2.1 *De se* and *de dicto*

The *de se-de dicto* distinction is made vivid by one of Lewis's examples:

Consider the case of the two gods. They inhabit a certain possible world, and they know exactly which world it is. Therefore, they know every proposition that is true at their world. Insofar as knowledge is a propositional attitude, they are omniscient. Still I can imagine them to suffer ignorance: neither knows which of the two he is. They are not exactly alike. One lives on top of the tallest mountain and throws down manna; the other lives on top of the coldest mountain and throws down thunderbolts. Neither one knows whether he lives on the tallest mountain or on the coldest mountain; nor whether he throws manna or thunderbolts. (Lewis, 1979, 139)

De dicto knowledge is knowledge about the way the world is. The gods have all the *de dicto* knowledge there is to have. *De se* knowledge is knowledge about who one is. Neither god has all the *de se* knowledge he could have, since neither god knows who he is.

Lewis starts from the idea that the content of an epistemic state is a set of possible worlds. To know a possible worlds proposition p is for every possible world compatible with what one knows to be contained in p . Manna and Thunder (as we shall call them) know which world they inhabit; for every possible worlds proposition p , Manna and Thunder know whether their world is in p or not. Still, both suffer ignorance, since neither knows who he is.

In order to model Manna and Thunder's ignorance, Lewis encourages us to re-think the nature of an agent's epistemic alternatives when it comes to *de se* knowledge and ignorance. Rather than taking them to be possible

worlds, Lewis suggests that we take them to be centered worlds instead. A centered world $\langle w', x' \rangle$ is compatible with what an agent x knows just in case, for all x can tell, she might be x' in w' . An agent x knows (*de se*) that she is F just in case every $\langle w', x' \rangle$ compatible with what she knows is such that x' is F in w' . To know a *centered proposition* (set of centered worlds) p is for every centered world compatible with what one knows to be contained in p .

The situation of the two gods can now be thought of as follows: Suppose the gods live in world β . Since they know every true proposition, β is the only possible world compatible with what they know. But since neither god knows who he is, there are two centered worlds compatible with what each knows, namely $\langle \beta, \text{Manna} \rangle$ and $\langle \beta, \text{Thunder} \rangle$. Since neither god knows whether he is Manna or Thunder, both of those centered worlds are compatible with what each god knows.

Our characterization of Lewis's theory actually differs slightly from Lewis's own. A central theme of Lewis (1979) is that we can use sets of centered worlds to do all the work sets of possible worlds can do and more (though Lewis (1994, 320) refers to this as a “cheap trick”). Any content that could be characterized as a set of possible worlds $\{w : \phi(w)\}$ could instead be characterized as the set of centered worlds $\{\langle w, x \rangle : \phi(w)\}$. The distinction between *de se* and *de dicto* contents can then be made in terms of *types* of sets of centered worlds. Following (Egan, 2006, 107), we can say that a *de dicto* content is a *boring* set of centered worlds, where a set of centered worlds p is boring just in case for all possible worlds w , and individuals x and y in w , $\langle w, x \rangle \in p$ iff $\langle w, y \rangle \in p$. A boring centered proposition doesn't distinguish between world-mates, and so there is a 1-1 correspondence between possible worlds propositions and boring centered propositions. A centered proposition is *interesting* iff it is not boring. A *de se* content is an interesting centered proposition.

In terms of this terminology, we could formulate our account of the inside-outside distinction as follows:

Uncentered Imagination (boring)

The content of an imagining from the outside is a boring set of centered worlds.

For all imaginings I , I is an imagining from the outside iff I 's content is a boring set of centered worlds.

Centered Imagination (interesting)

The content of an imagining from the inside is an interesting set of

centered worlds.

For all imaginings I , I is an imagining from the inside iff I 's content is an interesting set of centered worlds.

I don't think a lot hangs on the difference between thinking of *de dicto* contents as boring centered propositions or thinking of them as possible worlds propositions. I will use both at different points in this dissertation, depending on which approach best suits my purposes.⁵

1.2.2 *De se* attitude reports

Lewis's account of *de se* attitudes serves as the basis for an attractive account of *de se* ascription. Consider the following case:⁶

Bleeding John

John is looking at himself in the mirror, but he doesn't realize that it is he who he sees. The arm of the man in the mirror is bleeding, i.e. John's arm is bleeding. Knowing that those around him are a helpful sort, John thinks to himself, *That guy will receive help*. But not believing that his own arm is bleeding, John doesn't believe that he himself will receive help, i.e. he doesn't think to himself, *I will receive help*.

And consider the following two sentences:

- (1) a. John expects to receive help.
- b. John expects that he will receive help.

The data, first observed by Morgan (1970), is that (1a) is unambiguously false in this situation, whereas (1b) has at least one reading on which it is true. Note that John lacks the *de se* expectation that he himself will receive help, but that he has a *de dicto* (or *de re*) expectation that the person who is in fact identical to him will receive help. This suggests that

⁵One reason for preferring the centered worlds vs. possible worlds characterization of the *de se-de dicto* distinction is that the attempt to characterize all contents using only centered worlds has been challenged recently by Nolan (2006), who points out that Lewis's official account is incompatible with the existence of certain *de se* desires (e.g. the wish that one had never existed). I suspect that Lewis's account can be defended from Nolan's objection, but I won't pursue this further.

⁶This is a modified version of a case in von Fintel (2005).

(1a) can only be read as reporting a *de se* expectation, whereas (1b) can be read as reporting either a *de se* or a *de dicto* expectation (the latter giving us a reading that is true in this situation).

According to a standard syntactic theory, the subject of (1a)'s lower clause is *PRO*, a phonologically null pronoun that is controlled by the subject of the higher clause.⁷ So the real structure of (1a) is:

John expects (*PRO*) to receive help.

So our observation about the Bleeding John case can be put by saying that *PRO* gives rise to unambiguously *de se* readings.

Why does subject-control *PRO* give rise to unambiguously *de se* readings? The analysis of attitude verbs and *PRO* that we adopt is inspired by Chierchia (1989) and von Stechow (2002, 2003).⁸ This is couched in a more general framework of type-theoretic intensional semantics, similar to the system in von Fintel and Heim (2004). In this framework, semantic values are given relative to a context c , an index i , and a contextually-determined variable assignment g . A context is a triple $\langle w_c, t_c, x_c \rangle$ where x_c is the speaker at time t_c in possible world w_c . An index is just a possible world w_i (for the sake of ease, we ignore the temporal features of the sentences in question).

The general idea of our semantics is that the complement clause of an attitude verb can express either a possible worlds proposition or a centered proposition.⁹ Centered propositions are generated when an object-language λ -binder occurs inside the scope of the attitude verb and binds an element in its complement clause. Such a λ -binder may bind overt pronouns, but we also assume it can bind *PRO*. Indeed, we assume that *PRO* is *obligatorily* bound by a λ -binder that occurs in the scope of the nearest attitude verb.¹⁰

⁷Not every gerundive phrase has *PRO* as subject; but standard diagnostics (e.g. idiom and dummy-*it* tests) reveal that *imagines* is a subject-control, rather than a raising, verb.

⁸For a related, but slightly different, take on these issues, see Anand and Nevins (2004), von Fintel (2005), and Anand (2006).

⁹Note that this does not necessarily mean that attitude verbs are ambiguous. For example, $[\![\text{Sam believes}]\!]_{c,i,g}$ might denote a function whose domain is the union of the set of possible worlds propositions and the set of centered propositions. (The double brackets “ $[\!]$ ” denote a four-place interpretation function that takes expression-context-index-assignment quadruples to truth values.)

¹⁰This assumption can be spelled out in the compositional semantics by specifying the right phi-features for *PRO*; see von Stechow (2002, 2003) for details. For background on the relevant system of variable binding, see Heim and Kratzer (1998).

With these assumptions in place, we have the following semantic value (intension) for *PRO to receive help*:

$$\begin{aligned} & \lambda w. [\lambda_1 \text{ PRO}_1 \text{ to receive help}]^{c,w,g} \\ &= \lambda w. \lambda x. [\text{PRO}_1 \text{ to receive help}]^{c,w,g^{x/1}} \\ &= \lambda w. \lambda x. x \text{ receives help in } w. \end{aligned}$$

Note that this is (equivalent to) a centered proposition.¹¹ We assume the attitude verb *expects* shifts the index, so that we get the following truth condition for (1a):

$[\text{John expects } \lambda_1 \text{ PRO}_1 \text{ to receive help}]^{c,w_i,g} = 1$ iff all the $\langle w, x \rangle$ compatible with what John expects in w_i are such that x receives help in w .

So (1a) comes out false in Bleeding John, because it is not true in all the $\langle w, x \rangle$ compatible with what John expects that x receives help in w , since John lacks the relevant *de se* expectation.

The complement of *expects* on the intended reading of (1b), however, is a possible worlds proposition:

$$\begin{aligned} & \lambda w. [\text{he}_1 \text{ will receive help}]^{c,w,g} \\ &= \lambda w. g(\text{he}_1) \text{ receives help in } w. \end{aligned}$$

Note that no object language λ -binder occurs in the structure, and so he_1 is free; we assume the variable assignment g assigns John to he_1 . So (1b) will be true just in case John stands in the relation of expectation to the set of possible worlds in which John receives help. This gives us the following truth-condition for (1b):

$[\text{John expects that } \text{he}_1 \text{ will receive help}]^{c,w_i,g} = 1$ iff all the w compatible with what John expects in w_i are such that John receives help in w .

The reason (1b) comes out true in Bleeding John is that, on the relevant reading, that sentence is true just in case in all the w compatible with what John expects, John receives help in w . Since John *does* expect that the

¹¹Sometimes we characterize contents as sets of possible or centered worlds, sometimes as the characteristic function of such sets. The formulations are essentially equivalent, and we use both.

person who is in fact identical to John will receive help, (1b) is true. Thus, this semantic theory gets the desired result: (1b) has a true reading in this scenario, while (1a) does not.

I said that (1b) also has a reading on which it reports a *de se* expectation. Unlike (1a), (1b) has two readings. How is this predicted by our semantics? Assume that overt pronouns can, in certain circumstances, be bound by an object language λ -binder that occurs within the scope of the nearest attitude verb.¹² This means that there are two possible LFs for (1b):

- John expects that $\lambda_1 \text{ } he_1$ will receive help. (*de se*)
- John expects that he_1 will receive help. (*de dicto*)

The *de se* reading of (1b) will be generated when a λ -binder is present, since the intension of $\lambda_1 \text{ } he_1 \text{ will receive help}$ is a centered proposition. The *de dicto* reading will be generated when no such λ -binder is present, since, as we saw above, the intension of $he_1 \text{ will receive help}$ is a possible worlds proposition.

1.2.3 Imagination reports

We now have an account of *PRO* and attitude reports which has been motivated independently of any claims about the meaning of imagination reports, and we can now apply this account to the case of imagination.

I begin with an argument for Centered Imagination. When I introduced the inside-outside distinction I mentioned that it seems to correspond to a linguistic distinction. And often when philosophers write about the distinction between imagining from the inside and imagining from the outside, they emphasize that it is the distinction between imagining *doing* or *experiencing* something vs. imagining *that* one is doing or experiencing something. For example, consider this passage from Walton (1990, 29):

¹²The ‘in certain circumstances’ qualification is important. For consider sentence (a):

- (a) I expect that he will receive help.

Obviously this cannot be used to report a *de se* expectation. The constraint appears to be morphological: a pronoun can only be read *de se* if its phi-features agree with the phi-features of the subject of the attitude verb (e.g. *John expects that he will receive help* has a *de se* reading because the matrix subject and the subject of the complement clause exhibit phi-featural agreement). For discussion of this, see von Stechow (2002, 2003) and Schlenker (2003).

Imagining from the inside is... a form of self-imagining characteristically described as imagining *doing* or *experiencing* something (or *being* a certain way), as opposed to merely imagining *that* one does or experiences something or possesses a certain property.

This suggests that the standard way to report episodes of imagining from the inside is to use sentences like (2a), and that in order to report an episode of imagining from the outside one must use something like (2b):

- (2) a. Bernie imagined skiing down a steep hill.
- b. Bernie imagined that he was skiing down a steep hill.

Now according to the syntactic theory we are relying on, the real structure of (2a) also has *PRO* as the subject of the lower clause:

Bernie imagined (*PRO*) skiing down a steep hill.

The semantics we developed above entails that the semantic value (intension) of *PRO skiing down the hill* is the following object:

$$\begin{aligned} & \lambda w. [\lambda_1 \text{PRO}_1 \text{ skiing down a steep hill}]^{c,w,g} \\ &= \lambda w. \lambda x. [\text{PRO}_1 \text{ skiing down a steep hill}]^{c,w,g^{x/1}} \\ &= \lambda w. \lambda x. x \text{ is skiing down a steep hill in } w. \end{aligned}$$

This is (equivalent to) a centered proposition. So (2a) is true iff Bernie imagines the centered proposition $\{\langle w, x \rangle : x \text{ is skiing down a steep hill in } w\}$. And, as Walton observed, Bernie imagines from the inside skiing down a steep hill iff (2a) is true. So it follows that Bernie imagines from the inside skiing down a steep hill iff Bernie imagines the centered proposition $\{\langle w, x \rangle : x \text{ is skiing down a steep hill in } w\}$, which is an instance of **Centered Imagination**. Since there's nothing special about this particular case, we assume the point generalizes.

That's the argument for **Centered Imagination**. Now I'll argue for **Uncentered Imagination** by showing that, together with **Centered Imagination** and our semantic theory, it helps explain some facts about sentences like (2b):

- (2b) Bernie imagined that he was skiing down a steep hill.

We noted above that sentences like (2b) can be used to report outside imaginings. But note they can also be used to report inside imaginings; they have two readings. This seems to parallel the fact that (1b) has both a *de dicto* and a *de se* reading. If we assume our account of the inside-outside distinction, and assume the semantic theory sketched above, we can explain why (2b) has these two readings.

Our semantic theory entails that there are two LFs for (2b) (recall our discussion of (1b)):

- Bernie imagined that $\lambda_1 \ he_1$ was skiing down a steep hill.
- Bernie imagined that he_1 was skiing down a steep hill.

In the first LF, the overt pronoun he_1 is bound by a λ -binder occurring in the scope of the attitude verb. Thus, the intension of the complement clause of that LF is a centered proposition. If **Centered Imagination** is true, then we predict that (2b) can be read as reporting an inside imagining. On the second LF, the overt pronoun he_1 is free and no λ -binder occurs within the scope of the attitude verb. Thus, the intension of the complement clause of that LF is a possible worlds proposition. If **Uncentered Imagination** is true, then we predict that (2b) can be read as reporting an outside imagining. Thus, the combination of **Centered Imagination**, **Uncentered Imagination**, and our semantic theory explains why (2b) has both an outside and an inside reading.

Before moving on, I should mention that, on the proposed semantics, the first-person singular pronoun *I* is treated as just another pronoun, albeit with a distinctive presuppositional meaning. The lexical entry for *I* is:

$$[I_1]^{c,i,g} : g(I_1) = x_c. \ g(I_1)$$

This says that the semantic interpretation function is partial; it is not defined if $g(I_1) \neq x_c$. Where defined it maps I_1 to $g(I_1)$. (Note the slight difference from the standard Kaplanian entry: $[I]^{c,i,g} = x_c$.)

Like any other pronoun, *I* gets its value from the variable assignment; however, it presupposes that its referent is the speaker of the context. But the important thing is that this allows *I* to be bound, so that a sentence like (3) can be used to report an inside imagining:

- (3) I'm imagining that I'm skiing down a steep hill.

The inside reading results when *I* is bound; the outside reading results when *I* is free. In what follows, I will occasionally appeal to this idea.¹³

That completes our case for Centered and Uncentered Imagination. Although I will spend the rest of this chapter looking at a problem for this account, I think this account is adequate for many purposes. Indeed the issues discussed in the rest of Chapters 1, 2, and 3 may not be relevant for certain philosophical uses one might make of this account. For that reason, when I come to discuss the role these two forms of imagining play in discussions of personal identity, I will use the account of the inside-outside distinction developed above.

1.3 The problem of the ‘impersonal-*de se*’

In spite of this linguistic argument in favor of our account of the inside-outside distinction, there is a potential problem for it. The problem is that there seem to be cases of imagining from the outside that are also, in some sense, *de se*.¹⁴

Consider John Perry’s character, Rudolph Lingens (Perry, 1977, 21-22). Lingens is an amnesiac, who, in some intuitive sense, ‘doesn’t know who he is’. He is lost in the Stanford Library, and, we shall suppose, has stumbled upon a book entitled “*The Life of Rudolph Lingens*”. As Lingens reads the book, he learns more and more facts about Rudolph Lingens. But he’s still in the dark about who he himself is—he doesn’t realize that he is the man he is reading about. Let us suppose that Lingens puts the book down momentarily and begins to daydream.

We can distinguish two possible imaginings Lingens might have:

Situation 1

Lingens imagines *from the outside* that he himself is skiing down a steep hill. Someone asks him what he’s doing and he replies,

¹³ Again, binding can only take place in the appropriate linguistic context; see footnote 12. That the first person singular pronoun can be bound is supported by examples like (a):

- (a) Only I did my homework.

On one reading of (a) it says that for all *x*, if *x* is not the speaker, then *x* didn’t do *x*’s homework. It’s difficult to see how to generate this reading if *my* is not bound by *only I*. (This example is attributed to Irene Heim in Schlenker (2004)).

¹⁴ Thanks to Sarah Moss and Bob Stalnaker for pressing me to consider this problem.

(3) I'm imagining that I'm skiing down a steep hill.

Situation 2

Lingens imagines from the outside that the person he was just reading about, namely Rudolph Lingens, is skiing down a steep hill. Someone asks him what he's doing and he replies,

(4) I'm imagining that Rudolph Lingens is skiing down a steep hill.

Lingens doesn't know who he is. In his imagining in Situation 1, Lingens 'sees' someone skiing down a steep hill, and he stipulates that that skier is *him* ("Look at me go!" says Lingens). Intuitively, then, there is a difference between the content of the imaginings in Situations 1 and 2. And, moreover, it would seem that the imagining in Situation 2 is a *de dicto* imagining, whereas the imagining in Situation 1 is, in some sense, *de se*. But, importantly, *both* imaginings are from the outside.

If the imagining in Situation 1 is from the outside, then **Uncentered Imagination** tells us its content is a set of possible worlds. Which set? Presumably, the set of possible worlds in which Lingens is skiing down the hill. But then the imagining in Situation 1 has the same content as the imagining in Situation 2, which seems wrong.

If the imagining in Situation 1 is a *de se* imagining, then presumably it has a centered content (given that we've adopted Lewis's theory of the *de se*). But then **Centered Imagination** tells us that it is an imagining from the inside, which it is not. So the example causes trouble for both clauses of our theory of imagining.

I think the most natural response to the example is to give up the claim that *only* inside imaginings have centered contents. The idea would be that there are simply *two* kinds of *de se* imagining, inside imagining and what we might call *impersonal-de se imagining*. Lingens's imagining in Situation 1 is an example of the impersonal-*de se*.¹⁵

If impersonal-*de se* imaginings have centered contents, then presumably what Lingens imagines in Situation 1 is the set of centered worlds in which

¹⁵It might be more natural to call this second kind of imagining '*de se* imagining from the outside', but I don't want to do this because, as we will see in a moment, an imagining can be both from the inside and impersonal-*de se*. I want to avoid the terminological confusion that would result if we said that an imagining could be both from the inside and from the outside.

the center is skiing down a steep hill. If we adopted this view, we could still maintain that every imagining from the inside has a centered content. The revision would be to add to that that some imaginings from the outside – namely impersonal-*de se* imaginings – also have centered contents. Other outside imaginings would have uncentered contents. Here's the idea:

Alternative View

- (a) If x imagines that he is F from the inside, then all the $\langle w', x' \rangle$ compatible with what he imagines are such that x' is F in w' .
- (b) If x imagines ‘impersonally-*de se*’ that he is F , then all the $\langle w', x' \rangle$ compatible with what he imagines are such that x' is F in w' .
- (c) If x imagines something from the outside, and does not imagine something in the ‘impersonal-*de se*’ way, then the content of x 's imagining is a set of possible worlds.

Now if we accept the Alternative View, the inside-outside distinction cross-cuts that *de se-de dicto* distinction, and we simply need to find another way to characterize the inside-outside distinction. The Alternative View isn't an account of this inside-outside distinction, but rather a claim about how the two distinctions are related. Although the Alternative View doesn't give us an account of the inside-outside distinction, you might think it at least has the virtue of adequately characterizing the content of Lingens's imagining in Situation 1, something Uncentered and Centered Imagination do not do. But there is another type of case that the Alternative View cannot handle.

The main problem for the Alternative View is that an imagining can be both from the inside and impersonal-*de se*. This is seen most vividly when you combine Lingens-style ignorance cases with the *Lakoff phenomenon*. The Lakoff phenomenon is illustrated by these sentences (both inspired by Lakoff (1972)):

- (5) a. I imagined that I was Brigitte Bardot and that I kissed me.
- b. I imagined that I was Jesus and that I forgave me all my sins.¹⁶

It seems relatively clear what (5a) is saying: I imagine from the inside being Bardot, and I imagine kissing the person who is in fact me. (Sentence (5a)

¹⁶The original sentence in Lakoff (1972) is *I dreamt that I was Brigitte Bardot and that I kissed me*. The ‘dream’ version of (5b) appears in Heim (1994b).

1. Imagination and the *De Se*

may also have a reading on which it says that I'm imagining being Bardot and engaging in an act of self-kissing, but if it does, I wish to ignore that reading. The reading we're interested in is one in which the kisser and the 'kissee' are distinct.)

To see the problem this creates for the Alternative View, consider Situation 3:

Situation 3

Lingens imagines that he is Brigitte Bardot and that he is kissing the person who he in fact is. That is, someone asks him what he's doing and he replies by uttering the Lakoff sentence:

- (5a) I'm imagining that I am Brigitte Bardot and that I am kissing me.

Now I take it that the way Lingens imagines himself-*qua*-kissee in Situation 3 has something important in common with the way he imagines himself in Situation 1, when he imagines from the outside that he is skiing down a steep hill. In both cases, he 'sees' someone from an external point of view, someone he identifies as himself. When he imagines himself this way – whether the imagining is from the inside or from the outside – his imagining is impersonal-*de se*. So both inside and outside imaginings can be impersonal-*de se*.

The Alternative View runs into trouble with the imagining in Situation 3. To see the problem, first note that, in Situation 3, Lingens is imagining from the inside being Bardot and kissing exactly one person. Given clause (a) of the Alternative View, it follows that:

all the $\langle w, x \rangle$ compatible with what Lingens imagines are such that x is Brigitte Bardot in w and x is kissing exactly one person in w .

But Lingens is also imagining, in the impersonal-*de se* way, that he is being kissed by Bardot. Given clause (b) of the Alternative View, it follows that:

all the $\langle w, x \rangle$ compatible with what Lingens imagines are such that x is being kissed by Brigitte Bardot in w .

From those two claims it follows that:

all the $\langle w, x \rangle$ compatible with what Lingens imagines are such that x is Brigitte Bardot in w and x is kissing x in w .

Thus, the imagining is predicted to be one in which Lingens is imagining being Bardot engaged in act of self-kissing. But this is *not* what Lingens is imagining, and so the Alternative View gets the content of his imagining wrong.

We can construct a single case that is problematic for both our initial view (Centered and Uncentered Imagination) and the Alternative View:

Situation 4

Amnesiac Lingens takes himself to be a gentle soul. After reading of some of Rudolph Lingens's exploits, he comes to believe that this Rudolph Lingens fellow is a vicious character. He concludes that he is not Rudolph Lingens. He begins to wonder what it would be like to meet the terrible Lingens. He worries that Lingens might attack him if they were to meet, and he imagines this meeting. And in an effort to understand the mindset of Rudolph Lingens, he imagines the attack from Lingens's point of view. He might report his imagining by saying:

- (6) I imagined that I was Rudolph Lingens and that I attacked me.

Neither our initial theory nor the Alternative View correctly characterizes the content of this imagining. We can put the problem in linguistic terms. There are only two possibilities for the semantic value of *me* in (6): either it refers to the speaker (Lingens), or it marks the center (i.e. it is bound by a quantifier over centered worlds). If our original view was right, it would refer to Lingens. If the Alternative View was right, it would mark the center. But in this particular case, the difference between these views collapses, since the center *is* Lingens. Thus *both* of these views entail that the imagining will have the following as its content:

$$\{\langle w, x \rangle : x \text{ is Rudolph Lingens and } x \text{ attacks } x \text{ in } w\}$$

But this is not what Lingens is imagining: in his imaginative scenario, the attacker is distinct from the one attacked. So neither view is adequate.

1.4 Towards a solution

When you combine the Lakoff-phenomenon with the amnesiac-type cases, it begins to seem that, in the case of attitudes like imagining, there are

1. Imagination and the *De Se*

two different ‘self-concepts’ in play. Let me try to give an intuitive (and somewhat metaphorical) description of what’s happening in that sort of case. Suppose I were an amnesiac like Lingens. From my point of view, when I engage in a Lakoff-style imagining, my imaginary scenario contains two ‘selves’: there is my *counterfactual-self*, or the individual I am imagining being, and there is my *belief-self*, or the individual I believe myself to be. In Lakoff cases, these two selves come apart. In the Bardot-case, for example, my counterfactual self is Bardot, and my belief-self is the kissee. Note that, in this sense, the belief-self need not be identical to Dilip, since I may not know who I am, and may even have false beliefs about who I am.

One way to try to handle this sort of case would be to represent each of these ‘selves’ or ‘self-concepts’ with a different ‘center’: rather than characterizing imagination-alternatives as centered worlds, perhaps we should represent them with *doubly-centered worlds* $\langle w, x_1, x_2 \rangle$, with x_1 keyed to my counterfactual-self, and x_2 keyed to my belief-self.¹⁷

The idea would be that Lingens’s imagining in Situation 3 could be represented as follows:

$$\{\langle w, x_1, x_2 \rangle : x_1 \text{ is Brigitte Bardot in } w \text{ and } x_1 \text{ is kissing } x_2 \text{ in } w\}$$

Lingens’s imagining in Situation 4 could be represented in a similar manner:

$$\{\langle w, x_1, x_2 \rangle : x_1 \text{ is Rudolph Lingens and } x_1 \text{ attacks } x_2 \text{ in } w\}$$

The idea is that the first center, x_1 , is the ‘inside-center’, and the second center, x_2 , is the ‘impersonal-*de se*-center’.

Now we could proceed from here by formulating and exploring a theory based on this idea. The resulting theory would, I believe, be adequate to cover the cases we’ve discussed, and would give us an account of the inside-outside distinction. But it might, I fear, look a bit *ad hoc*. Instead of building directly on this idea, I want to turn to a topic which should seem at least somewhat relevant to the present discussion: how, in general, we should handle attitudes about individuals, i.e. *de re* attitudes. What I will argue in the following two chapters is that there are two plausible approaches to *de re* attitudes in which a theory of *de se* imagining that employs the ‘two centers’ idea embeds comfortably. I think looking more carefully at how we might treat *de re* attitudes will make this idea seem more plausible.

¹⁷Steve Yablo first suggested this idea to me.

1.4.1 Interlude on aspects of the inside-outside distinction

Before moving on to the next chapter, I briefly want to note some interesting aspects of the inside-outside distinction which I will unfortunately not be discussing in any detail.

One thing the Lakoff phenomenon makes vivid is that attitudes other than imagining seem to admit of an inside-outside distinction, as Lakoff (1972) notes. Lakoff's original example involved dreaming, not imagining: *I dreamt that I was Brigitte Bardot and that I kissed me*. Desiring and wishing are similar, as Lakoff also observed: I may want to be President but not want myself to be President (I want to be President, but think the person I in fact am would do a terrible job, and do not want *that person* to be President). Another example: Suppose that Sam is my enemy. If I am a certain type of self-loathing character, I may want to be Sam and beat me up; I want to be Sam and beat up the person I in fact am. What these attitudes all seem to have in common is that they are *counterfactual* in the sense that it is rational to bear these attitudes towards a content *p* even if one believes that *p* is false. (Part of the next chapter will be devoted to exploring counterfactual *de re* attitudes.)

Memory provides an interesting case.¹⁸ As has often been observed in the literature, remembering admits of an inside-outside distinction, a fact that also seems to be reflected by the grammar:

- (7) a. I remember delivering the speech.
b. I remember myself delivering the speech.

Here (7b) may report a memory from the outside: I 'see' the speaker – me – from a point of view that is not the speaker's. (7a), on the other hand, means I have a memory of giving the speech from the inside. If I was the only one who delivered the speech, then, while others can remember me delivering the speech, only *I* can remember *delivering* the speech.¹⁹ But, interestingly, remembering seems not to be a counterfactual attitude.

Do Lakoff sentences involving *remembers* make sense?

- (8) I remember being Brigitte Bardot and kissing me.

¹⁸See Higginbotham (2003) for discussion.

¹⁹There is also a difference between (7a) and (7b) on the one hand, and *I remember that I delivered the speech*, on the other. The latter can be true even if I have do not remember the *event* of my giving the speech at all, but simply know that I in fact did give it (perhaps because I am told by others that I gave it). This seems related to the psychologists' distinction between *semantic* and *episodic* memory.

1. Imagination and the *De Se*

I *think* I can make sense of this: I was once Brigitte Bardot, and when I was, I kissed the person that I am now. If this does make sense, it would be interesting to see what (if anything) this means for the theory of personal identity over time.

Before wrapping up this interlude, let me also point out that, within the category of imagining, we get something analogous to the inside-outside distinction with respect to time. I can imagine that it's now 1942: the temporal perspective of my imagining is set at 1942. I can also imagine that it's 50 years from now, i.e. we are 50 years in the future from the present time (whatever that time in fact is; I can imagine this even if I don't know what year it is now). The two uses of *now* in those sentences seem to parallel the two uses of *I/me* in the Lakoff sentences.

As you can see, our discussion thus far has only scratched the surface of these issues. A full account of the inside-outside distinction should have something to say about them all. Unfortunately, we must leave these as matters for future inquiry.

Chapter 2

Imagination, Descriptivism, and the *De Re*

In this chapter and the next, I discuss two related issues: what to say about *de re* attitudes if one accepts Lewis's view of the *de se*, and how to solve the problem of the impersonal-*de se*. (Just how these issues are related will gradually emerge.) I begin by considering Lewis's own approach to *de re* attitudes, which is a 'descriptivist' account of the *de re*. In this chapter, I accept descriptivism about the *de re*, but argue that Lewis's account, as formulated, cannot be extended to counterfactual *de re* attitudes (e.g. imagining, dreaming, wishing). I argue that a solution to this problem can be found if we accept the idea that counterfactual attitudes have a *two-dimensional* structure: their contents are *two-dimensional centered intensions*, or functions from centered worlds to centered propositions. I show how this hypothesis solves both the problem of counterfactual attitudes and the problem of the impersonal-*de se*.

One reason for being interested in Lewis's account of the *de re* is that it has given rise to an account of attitude ascriptions influential in the formal semantics literature; this chapter ends with a discussion of attitude ascriptions.

Chapter 3 looks at these same issues through an anti-descriptivist lens.

De re attitudes are thoughts about individuals. Many philosophers think that to have a *de re* attitude about something one must stand in a *relation of acquaintance* to that thing. In order for me to have beliefs, desires, etc. about a certain individual, there must be a causal link between me and that individual, a link through which information about that object can travel. A relation of acquaintance is a causal link of this sort. Lewis (1979, 155) puts the idea this way:

I and the one of whom I have beliefs *de re* are so related that

there is an extensive causal dependence of my states upon his; and this causal dependence is of a sort apt for the reliable transmission of information.

This is a plausible idea, but just what counts as being acquainted with something is not entirely clear. Usually perceptual acquaintance with an individual is taken to be sufficient for one to have *de re* beliefs about it: if I see Ortcutt sneaking around on the docks, then I can come to have *de re* beliefs about Ortcutt. On the other hand, being related to the *res* by ‘pure description’ is usually taken to be insufficient: I think, on purely general grounds, that the shortest spy is under six feet tall. Assuming that there is a uniquely shortest spy, I still don’t count as having a *de re* belief about him or her, since (we can suppose) I do not stand in an appropriate causal relation to him or her.

In an effort to clarify what he has in mind, Lewis offers a list of examples of possible types of *de re* belief (Lewis, 1979, 154-155). On Lewis’s view, I can have *de re* beliefs:

- about my acquaintances, present or absent;
- about contemporary public figures prominent in the news;
- about the famous dead who feature prominently in history;
- about authors whose works I have read;
- about strangers now face to face with me;
- about strangers I am somehow tracing, such as the driver of the car ahead of me, or the spy I am about to catch because he left so many traces; and
- about people I have heard about under a particular name, e.g. about the one I have heard of under the name “David Hume”.

I’ll assume that the notion of a relation of acquaintance is clear enough for present purposes, and won’t seek to clarify it any further here.¹

¹For more discussion of this issue, see Burge (1977), Lewis (1979, 153 - 155), Stalnaker (1988, 164 - 65), and Pryor (2004).

2.1 The Naive Theory

In the previous chapter, we were tacitly assuming something like the following account of *de re* imagination:

Naive Theory

x imagines (*de re*) in w that y is F iff:

- i. there is an acquaintance relation R such that x bears R to y in w , and
- ii. all the $\langle w', x' \rangle$ compatible with what x imagines in w are such that y is F in w .

(I take it that (ii) can only be true if (i) is true, in which case (i) is redundant.) We could, of course, extend this account to attitudes other than imagining, most notably belief. In discussing the Naive Theory, I will focus on its account of *de re* belief. There are at least two objections to this theory (Lewis, 1981).

The first is that it runs into trouble with some well-known puzzle cases, cases which have dominated the study of *de re* attitudes. Quine (1956) tells a story in which Ralph believes that Bernard J. Ortcutt, the town mayor and a pillar of the community, would never engage in espionage. But Ralph believes that *that fellow* [we point at a shadowy figure in a trench coat] is a spy. Unbeknownst to Ralph, it turns out that that fellow *is* Ortcutt the mayor.

The Naive Theory has the consequence that Ralph has inconsistent beliefs. But if rationality requires that one's beliefs be consistent, then the Naive Theory has the consequence that Ralph is not rational. But that seems wrong: Ralph lacks information, not logical acumen. Faced with such cases, some philosophers give up the idea that rationality requires consistency, but we shall be exploring the possibility of avoiding that result.²

Additionally, since no centered worlds are compatible with what Ralph believes, it would seem that Ralph does not have a coherent conception of what the world is like. But this too seems wrong: Ralph *does* have a

²One view is that one can rationally hold inconsistent beliefs, as long as the inconsistent propositions one believes are presented under different ‘guises’ or ‘modes of presentation’. It may turn out that this is the best we can do. But appealing to the primitive notion of a ‘guise’ is not illuminating, and it is at least worth trying to see whether we can do better.

coherent conception of what the world is like. Again, what he lacks is a certain piece of information, namely that the man he saw giving the rousing speech is the man he saw sneaking around in a trench coat: in the world according to Ralph, there is a mayor named “Ortcutt” and a spy in trenchcoat, and the mayor is not the spy.

The second problem arises if individuals have non-trivial essential properties. Suppose Ortcutt essentially has DNA sequence d . Then every possible world w in which Ortcutt exists is one in which Ortcutt has DNA sequence d . According to the Naive Theory, if Ralph believes anything about Ortcutt, Ortcutt exists in every one of Ralph’s centered belief worlds. Given the essentiality of DNA sequences, it follows that Ortcutt has DNA sequence d in every one of Ralph’s centered belief worlds, and from this it follows that Ralph believes that Ortcutt has DNA sequence d . But this just seems wrong—it seems that Ralph could have beliefs about Ortcutt without having *any* opinions about Ortcutt’s DNA.

For these two reasons, Lewis rejects the Naive Theory. We shall follow him in this.

2.2 Lewis on *de re* attitudes

Lewis’s account of *de re* attitudes is parasitic on his account of *de se* attitudes. In a sense, Lewis doesn’t think the *de re* constitutes a distinctive class of attitudes: *de re* attitude *ascriptions* report facts that are partly psychological and partly about the attitude holder’s relation to his environment. The psychological part is simply a *de dicto* or a *de se* attitude.³

Here is the account:

Doxastic Lewis

x believes (*de re*) in w that y is F iff there is an acquaintance relation R such that:

- i. x bears R uniquely to y in w , and
- ii. x believes (*de se*) in w that the unique thing to which she bears R is F . (That is: all the $\langle w', x' \rangle$ compatible with what x believes in w are such that there is a y' in w' such that x' bears R uniquely to y' in w' and y' is F in w').

³The idea of factoring the *de re* into a psychological component and a fact about the attitude holder’s relation to the world is due to Kaplan (1968). More on this later.

2. Imagination, Descriptivism, and the *De Re*

This formulation makes explicit the way in which Lewis's account 'factorizes' a *de re* belief into a psychological component which is purely *de se*, and a non-psychological component concerning the agent's relation to the *res*.

If Ralph has a belief in w about Ortcutt in virtue of bearing R uniquely to him, then any of the centered worlds $\langle w', x' \rangle$ compatible with what Ralph believes contain a unique individual y' to which x' bears R in w' . In the terminology of Lewis (1983a), y' for x' in w' is a *counterpart by R-acquaintance* of Ortcutt for Ralph in w . In other words, y' is 'Ortcutt's representative' in Ralph's centered belief world $\langle w', x' \rangle$, or Ortcutt's representative *relative to R* in $\langle w', x' \rangle$.

In the puzzle case, Ortcutt has *two* different representatives in any one of Ralph's centered belief worlds. Suppose Ralph bears two different relations of acquaintance to Ortcutt: he knows Ortcutt as the mayor, because he heard the man give a rousing speech on the stump; and he knows Ortcutt as the shadowy figure in a trenchcoat because he saw the man in a trenchcoat in such-and-such circumstances. Ralph believes that Ortcutt is not a spy relative to the first relation of acquaintance, but he believes that Ortcutt is a spy relative to the second one. On Lewis's account, this means that: (i) Ralph heard Ortcutt give a rousing speech on the stump, and Ralph believes (*de se*) that the man he heard giving a rousing speech on the stump is not a spy; and (ii) Ralph saw Ortcutt in a trenchcoat in such-and-such circumstances, and Ralph believes (*de se*) that the man he saw in a trenchcoat in such-and-such circumstances is a spy. Once one is clear about Ralph's psychology, no puzzle remains.

Thus, any centered world $\langle w, x \rangle$ compatible with what Ralph believes contains two individuals, y and z such that x saw y give a rousing speech in w , and x saw z sneaking around in a trenchcoat in w . z is a spy in w , y is not. y is Ortcutt's representative (relative to the 'speech' acquaintance relation) in $\langle w, x \rangle$, and z is Ortcutt's representative (relative to the 'trenchcoat' acquaintance relation) in $\langle w, x \rangle$.⁴

The Ortcutt case is a case in which someone thinks one is two, in virtue of being acquainted with the same thing in two different ways. Also familiar are cases of mistaken identity, or thinking that two are one. Suppose a charming dog, Sparky, runs into my office on Tuesday. Then on Friday, a similar-looking dog, Barky, runs into my office. I think, *Oh, here's that friendly dog again*. But I am mistaken: Sparky is not Barky. I have a

⁴See Lewis (1994, 323) for a similar treatment of Kripke's (1979) case of puzzled Pierre.

'doubly *de re*' belief about Sparky and Barky (in virtue of seeing Sparky run into my office on Tuesday and seeing Barky run into my office on Friday) to the effect that the former is the latter. On Lewis's account, my *de re* belief can be decomposed into the following three facts:

- Sparky is the unique dog I saw in my office on Tuesday;
- Barky is the unique dog I saw in my office on Friday; and
- all the $\langle w, x \rangle$ compatible with what I believe are such that the unique dog that x sees in his office on Tuesday in w is the unique dog that x sees in his office on Friday in w .

Thus, the psychological part of my *de re* belief is a perfectly coherent *de se* belief. Again, once we are clear about the subject's psychology and his relation to the world, no puzzle remains.

Lewis's account also avoids the Naive Theory's problem with essential properties. Ralph bears R to Ortcutt, and has beliefs about Ortcutt, but has no opinion about Ortcutt's DNA sequence. All the centered worlds $\langle w, x \rangle$ compatible with what Ralph believes contain a unique individual y to whom x bears R . If Ralph has no opinion as to whether Ortcutt has DNA sequence d , then there is at least one $\langle w, x \rangle$ compatible with what Ralph believes in which the unique individual y to whom x bears R in w is such that y has DNA sequence d in w , and there is also at least one $\langle w', x' \rangle$ compatible with what Ralph believes in which the unique individual y' to whom x' bears R in w' is such that y' does *not* have DNA sequence d in w' . If one's DNA sequence is essential to one, then $y \neq y'$. On Lewis's view, the individual who represents Ortcutt in a given one of Ralph's centered beliefs worlds need not be identical to Ortcutt, nor need he be identical to the individual who represents Ortcutt in any of the other centered belief worlds.⁵

It's worth noting that Lewis's view is a variation on a theme that traces back to Quine (1956) and Kaplan (1968). The basic idea of this approach is to reduce the problematic *de re* attitudes to the less problematic *de dicto* (or in Lewis's case, *de se*) attitudes. (*De re* attitudes are thought to be problematic since they give rise to the puzzles above.) Abstracting from the specific details of Kaplan's proposal, the general idea is that a *de re* ascription consists of two components: First, a non-psychological

⁵Essential properties do, however, create a slightly different problem for Lewis's account. See footnote 12 in §3.2.

fact about how the agent is related to the *res*—namely, that the two are ‘acquainted’. And second, an underlying *descriptive belief*. The underlying descriptive belief must be of a certain form: assume that the agent has a *descriptive concept – the G* – under which she thinks of the *res*, and which the *res* satisfies (the *res* is the unique *G*). Then, if the agent believes *de re* that the *res* is *F*, the agent’s underlying *de dicto* (or *de se*) belief is the belief that the *G* is *F*.

This basic idea can be implemented in a number of ways. For example, in a possible worlds framework, the content of *x*’s underlying descriptive belief would be the set of possible worlds *w* in which the *G* in *w* is *F* in *w*. In that setup, the descriptive concept *the G* is a function from worlds to individuals. In Lewis’s retelling of this story, all belief contents are sets of centered worlds, and so the descriptive concept *the G* is a function from centered worlds to individuals. (In these modal frameworks, descriptive concepts are not linguistic descriptions, nor need they be expressible by a linguistic description; fundamentally, they are functions from possibilities to individuals.) Lewis also stipulates that the descriptive concept be determined by the relation of acquaintance that the agent bears uniquely to the *res*, so that the psychological and non-psychological components of a *de re* belief are systematically related.⁶ Thus, on Lewis’s picture, the descriptive concept *the G* is always something like *the one to whom I bear R uniquely*, for some acquaintance relation *R*. But despite these differences, it is clear that Lewis’s view is simply an instance of this broader Kaplanian idea. I mention this here, since the problem I am about to discuss applies not just to Lewis’s theory, but to the more general picture behind it.

2.3 The problem of counterfactual attitudes

So far we have only talked about *de re beliefs*. The problem I shall outline concerns not this account of belief, but rather its generalization to other attitudes. Here is what the generalized theory says, for *any* attitude *A*:

General Lewis

x A’s (*de re*) in *w* that *y* has *F* iff there is an acquaintance relation *R* such that:

- i. *x* bears *R* uniquely to *y* in *w*, and

⁶Though see Lewis (1994, 318-319) for a slightly looser account of how *de re* belief works.

- ii. $x \mathcal{A}'s (de\ se)$ in w that the unique thing to which she bears R is F . (That is: all the $\langle w', x' \rangle$ compatible with what $x \mathcal{A}'s$ in w are such that there is a y' in w' such that x' bears R uniquely to y' in w' and y' is F in w').

' \mathcal{A}' ' can be replaced by any attitude verb, e.g. *believes*, *knows*, *desires*, and, importantly for what follows, *wishes*, *dreams*, and *imagines*.

As far as I know, Lewis only offered his theory as a theory of *de re* belief. So nothing he says is directly threatened by the problem presented here. Thus, when I say that Lewis's account is incompatible with certain apparent possibilities, I really mean that **General Lewis** is incompatible with certain apparent possibilities.

The problem for **General Lewis** comes in the form of a counterexample. Suppose Ralph sees Ortcutt sneaking around on the docks, and suppose this is the only way in which Ralph is acquainted with Ortcutt. Call the relation of acquaintance Ralph bears uniquely to Ortcutt ' R '. It seems that in virtue of bearing R uniquely to Ortcutt, Ralph is in a position to have various *de re* attitudes about Ortcutt: he might, for example, come to believe that Ortcutt is a spy, by believing that the individual to which he is uniquely R -related is a spy.

But it also seems that bearing R uniquely to Ortcutt puts Ralph in a position to think about various counterfactual situations in which Ortcutt figures. That is, Ralph might have *de re imaginings* about Ortcutt. Now the problem for **General Lewis**: it seems that Ralph could imagine a counterfactual scenario in which he is *not* acquainted with Ortcutt. For example, Ralph could imagine a scenario in which he and Ortcutt never cross paths, and do not know of each other's existence. Perhaps in Ralph's imaginary scenario, he lives in Tokyo and Ortcutt in Alaska, and their two lives are as causally separate as two lives could be. In this example, it seems true that Ralph imagines that he is not acquainted with Ortcutt. But on Lewis's account, this cannot be true.

According to **General Lewis**, for it to be true in a world w that Ralph imagines that he is not acquainted with Ortcutt, there must be an acquaintance relation R that has two properties:

- i. Ralph bears R uniquely to Ortcutt; and
- ii. All the centered worlds $\langle w, x \rangle$ compatible with what Ralph imagines are such that there is a y in w such that x bears R uniquely to y in w and x is not acquainted with y in w .

But no centered world $\langle w, x \rangle$ can meet the condition specified in (ii), given that R is an acquaintance relation. For x must both bear R uniquely to some y in w and, at the same time, not be acquainted with that same y in w . But if x bears R to y , then x is thereby acquainted with y . So no centered world is compatible with what Ralph imagines, and the content of Ralph's imagining is empty. Thus, General Lewis incorrectly predicts that Ralph cannot imagine that he is not acquainted with Ortcutt.

The problem is not limited to the attitude of imagining. Suppose Ralph sees a man kissing his wife; it is dark, so Ralph can see little else about the man. The man in question is Ortcutt. Ralph bears a relation of acquaintance R uniquely to Ortcutt: the relation x bears to y just in case x sees y kiss x 's wife. This puts him in a position to have *de re* attitudes about Ortcutt: for example, he might believe that Ortcutt is a man of low character. The trouble comes with a *de re wish* that Ralph might have about Ortcutt: he might wish that Ortcutt had not kissed his wife. Again, according to General Lewis, two things must be true for Ralph to so wish:

- i. Ortcutt is the man Ralph sees kiss his wife, and
- ii. All the centered worlds $\langle w, x \rangle$ compatible with what Ralph wishes are such that there is a unique y in w such that x sees y kiss x 's wife in w , and y does not kiss x 's wife in w .

But since seeing is a factive attitude, no centered world satisfies the condition set forth in (ii). General Lewis entails the falsehood that Ralph cannot, in this scenario, wish that Ortcutt did not kiss his wife.

It seems that the problem will arise for any type of *de re* attitude that one can reasonably bear to a content p even when one believes that p is false: it will arise for any *counterfactual attitude* (§1.4.1). We have the ability to represent individuals in counterfactual circumstances, circumstances that we believe do not obtain. This ability serves a variety of purposes, and is something that a theory of attitudes ought to accommodate.

As I noted earlier, Lewis's view of *de re* attitudes is the result of merging his view of *de se* attitudes with a broadly Kaplanian understanding of the *de re*. I also mentioned earlier that the problem of counterfactual attitudes is a problem not just for Lewis's view, but for the Kaplanian picture behind it. This is clearest when we look at an implementation of Kaplan's idea in a possible worlds framework (which is of course only minimally different from the centered worlds framework). Suppose one took the following view of *de re* attitudes:

$x \mathcal{A}'s (de re) in w$ that y has F iff there is an acquaintance relation R such that:

- i. x bears R uniquely to y in w ,
- ii. y is the unique G in w , and
- iii. all the worlds w' compatible with what $x \mathcal{A}'s$ in w are such that the G in w' is F in w' .⁷

It's not hard to see that this view will run into the problem of counterfactual attitudes: Suppose Ralph believes (*de re*) that Ortcutt is a spy. Suppose Ralph's descriptive concept of Ortcutt is *the bald man in the brown hat*, here understood to be the function which maps each world w to the unique bald brown-hatted man in w . Then the content of the relevant underlying belief is the set of worlds w' such that the bald man in the brown hat in w' is a spy in w' . But surely Ralph can also imagine (*de re*) that Ortcutt is a hirsute man with no hat. But the content of Ralph's underlying imagining would then be the set of worlds w' such that the bald man in the brown hat in w' is not bald and has no hat in w' . But that is the empty set.

A question remains about the scope of the problem. It's easiest (for me, at least) to see how the problem arises in these modal frameworks, but its application may be wider. Consider the following broad view:

$x \mathcal{A}'s (de re) in w$ that y has F (relative to acquaintance relation R) iff:

- i. x bears R uniquely to y in w ,
- ii. y is the unique G in w , and
- iii. $x \mathcal{A}'s (de dicto)$ in w that the G is F .

It seems to me that any specific proposal which is an instance of this schematic view will face the problem of counterfactual attitudes. For suppose that Ralph's descriptive concept of Ortcutt is *the G* (here we remain neutral on just how the theory interprets the notion of a descriptive concept). The problem is that any view of this sort will wind up entailing that Ralph cannot imagine/wish/dream (*de re*) that Ortcutt is not the G , for this would be to imagine/wish/dream (*de dicto*) that the G is not the

⁷One would probably want to require the descriptive concept *the G* to be sufficiently 'vivid' (cf. Kaplan, 1968). Alternatively, one might require that the agent's *de dicto* beliefs about the G be counterfactually sensitive to facts about y .

G. This seems like it would be a problem even for views not couched in a possible worlds or centered worlds framework. But, without specific views on the table, it is difficult to establish this general point. Since our main concern is what to say about the *de re* if one accepts Lewis's view of the *de se*, we leave as a question for future inquiry just how broadly the problem of counterfactual attitudes applies.⁸

2.4 A two-dimensional account

Suppose we want to maintain Lewis's view of *de re* belief. What should we say about counterfactual attitudes?

On Lewis's account, we think of individuals under descriptive concepts. For example, Ralph might think of Ortcutt under the description *the one I saw sneaking around on the docks*. The problem of counterfactual attitudes arises because Ralph can imagine situations in which Ortcutt fails to satisfy the description under which Ralph thinks of him. Let's try to get inside Ralph's mind for a moment. When Ralph imagines that he is not acquainted with Ortcutt, Ralph might put this to himself by saying, "The guy who is *in fact* the one I saw sneaking around on the docks—I'm imagining that I am not acquainted with *him*. Since I'm imagining that I'm not acquainted with him, he's not 'the one I see sneaking around on the docks' *in my imaginary scenario*. Indeed, I'm not imagining that I saw anyone sneaking around on the docks."

Now, it is very tempting to think that when Ralph says to himself "the guy who is in fact the one I saw sneaking around on the docks", Ralph 'reaches back' to the actual world and puts the the actual person that he saw sneaking around on the docks – namely, Ortcutt – into his centered imagination worlds, so that the content of his imagining is the set of centered worlds in which the center is not acquainted with Ortcutt. But this simply gets us back to the Naive Theory, and we know that that view faces troubles of its own.

Nevertheless, something seems right about this idea of 'reaching back'. But maybe when Ralph says to himself "the guy who is in fact the one I saw sneaking around on the docks", he reaches back not to the *actual world*, but to his *centered belief worlds*. He 'grabs' the person who satisfies that descriptive concept in his centered belief worlds and imagines that

⁸Perhaps if we replace "the *G*" in clause (iii) with "the actual *G*", the resulting theory will avoid the problem. But in any modal theory of attitudes, rigidifying the description just leads back to (some version of) the Naive Theory.

he is not acquainted with *that guy*. But that guy need not be the one the center of the centered *imagination* worlds saw sneaking around on the docks—indeed, there may not be anyone that the center of the centered imagination worlds sees sneaking around on the docks, for Ralph may not be imagining that he’s seeing someone sneaking around on the docks.

The idea here is that the content of Ralph’s imagining is somehow *anaphoric on* the content of his beliefs. I think this is basically the right intuition, but the way I put it in the above paragraph isn’t quite adequate, because of a difference between the actual world and Ralph’s centered belief worlds, namely that the former is one and the latter are many. If Ralph *were* to reach back to the actual world, he would be able to ‘grab’ the guy who he saw sneaking around on the docks in that world. The reason this makes sense is that there is only *one* actual world for Ralph to reach back to, and so there is a unique individual in that world who Ralph saw sneaking around on the docks. That individual could then be imported into Ralph’s centered imagination worlds. But the idea that Ralph reaches back to his centered belief worlds cannot work exactly like that. The problem is that, so long as Ralph is not maximally opinionated, there will be *more than one* centered world compatible with what he believes. And there is simply no guarantee that the individual who satisfies the description “the one I saw sneaking around on the docks” relative to one of Ralph’s centered belief worlds will be identical to the individual who satisfies that description relative to another. There might be different individuals in these different centered worlds, each of whom satisfies that description relative to his ‘own’ respective centered belief world. Which of these many individuals is the one we should import into Ralph’s centered imagination worlds?

Forget that question for the moment, and think about what we would say if Ralph *were* maximally opinionated. Suppose, for example, that $\langle w, x \rangle$ were the only centered world compatible with what Ralph believes, and that there is a unique individual – call him “Bill” – in w such that x bears R uniquely to Bill in w . Bill is Ortcutt’s representative in $\langle w, x \rangle$. In this case, if Ralph were to imagine that he was not acquainted with Ortcutt, it seems like it would be reasonable to say that the content of Ralph’s imagining would be the set of centered worlds $\langle w', x' \rangle$ in which x' is not acquainted with *Bill* in w' . Bill would count as Ortcutt’s representative in $\langle w', x' \rangle$ in virtue of the fact that he is Ortcutt’s representative in Ralph’s centered belief world $\langle w, x \rangle$.

We still face the problem about what to do if Ralph is not maximally opinionated. But the above case is suggestive. One thing it suggests is

that if we were to pick one of Ralph's centered belief worlds and then pretend momentarily that it was his *only* centered belief world, then we would be able to characterize his centered imagination worlds *relative to* that centered belief world. Note that we could go through each of Ralph's centered belief worlds and do this.

The emerging idea is that we can characterize *a set* of centered worlds compatible with what Ralph imagines *relative to* one of his centered belief worlds. What does this mean? The previous paragraph suggests a procedure for deciding whether or not a centered world $\langle w, x \rangle$ is compatible with what Ralph imagines relative to one of his centered belief worlds $\langle w', x' \rangle$. The procedure is simple: we simply ask ourselves, "Would $\langle w, x \rangle$ be compatible with what Ralph imagined if $\langle w', x' \rangle$ were his only centered belief world?" If the answer is 'yes', then $\langle w, x \rangle$ is compatible with what Ralph imagines relative to $\langle w', x' \rangle$. If the answer is 'no', then it is not.

This 'decision procedure' idea shouldn't be taken *too* seriously; it's really just a heuristic for getting an intuitive grip on what I think we should regard as a primitive notion in our theory of imagining: the notion of a centered world's being compatible with what someone imagines *relative to* one of her centered belief worlds. (For one thing, we may not always be that good at assessing counterfactuals of the form *If x were maximally opinionated, then $\langle w', x' \rangle$ would be compatible with what she imagines.*)

If this view is on the right track, then what it suggests is that the content of Ralph's imagining determines a function that takes a centered belief world $\langle w, x \rangle$ and returns a set of centered worlds, the set of centered worlds compatible with what Ralph imagines relative to $\langle w, x \rangle$. Formally, then, the content of Ralph's imagining determines a function from centered worlds to sets of centered worlds, or a *two-dimensional centered intension*.⁹ We shall identify the contents of imaginings with such intensions. On this view, "the set of centered worlds compatible with what Ralph imagines" is an improper definite description, for there is no such unique set. There are only sets of centered imagination worlds relative to centered belief worlds.

Although this view does not allow us to say what it is for a centered world to be compatible with what Ralph imagines *simpliciter* (we can only characterize this notion of compatibility relative to one of his centered belief worlds), we *can* say what it is for a *pair* of centered worlds to be compatible with what Ralph imagines *simpliciter*. This 'absolute' notion

⁹A two-dimensional centered intension can also be represented by a set of pairs of centered worlds, or by a function from centered worlds to a function from centered worlds to truth values.

can be defined in terms of the ‘relative’ notion as follows:

$\langle\langle w', x' \rangle, \langle w'', x'' \rangle \rangle$ is compatible with what x imagines in w iff:
 $\langle w', x' \rangle$ is compatible with what x believes in w , and $\langle w'', x'' \rangle$
is compatible with what x imagines in w relative to $\langle w', x' \rangle$.

Then we can say that x imagines a two-dimensional centered intension p iff all the $\langle\langle w', x' \rangle, \langle w'', x'' \rangle \rangle$ compatible with what x imagines are contained in p .

It should be clear how this account solves the problem of counterfactual attitudes. Suppose Ralph bears R to Ortcutt. When Ralph imagines that he is not acquainted with Ortcutt, we can characterize Ralph’s state of mind as follows: Any arbitrary centered world $\langle w, x \rangle$ compatible with what he believes contains a unique individual y to whom x is uniquely R -related; y is Ortcutt’s representative in $\langle w, x \rangle$. Relative to that centered belief world, the set of centered worlds $\langle w', x' \rangle$ compatible with what he imagines is such that x' is not acquainted with y in w' ; y represents Ortcutt throughout these imagination alternatives in virtue of representing him in the centered belief world $\langle w, x \rangle$. On this account, the content of Ralph’s imagining can be characterized by the following two-dimensional centered intension:

$\{ \langle\langle w, x \rangle, \langle w', x' \rangle \rangle : \text{there is a } y \text{ in } w \text{ such that } x \text{ is uniquely } R\text{-related to } y \text{ in } w, \text{ and } x' \text{ is not acquainted with } y \text{ in } w' \}$

Suppose Ralph thinks of Ortcutt under the descriptive concept *the G*. Our account allows us to first search one of Ralph’s centered belief worlds, find the individual who, in that centered belief world, is the unique G , and so represents Ortcutt there. We then use that individual to characterize Ortcutt’s imagination alternatives relative to the designated centered belief world. This procedure allows to identify that individual independently of the properties he has in the centered imagination worlds, which means that he need not be the thing (if there is one) that is the unique G in the centered imagination worlds. This is how we avoid the problem of counterfactual attitudes.

Our general account of *de re* imagining will look like this:

2D *De Re* Imagination

x imagines *de re* in w that y is F iff there is an acquaintance relation R such that:

- i. x bears R uniquely to y in w , and
- ii. every $\langle\langle w', x' \rangle, \langle w'', x'' \rangle \rangle$ compatible with what x imagines in w is such that there is a unique y' in w' such that x' bears R to y' in w' , and y' is F in w'' .

The first centered world in any pair of centered worlds compatible with what an agent imagines is one of the agent's centered belief worlds. Thus, on this account, to say, for example, that $\langle\langle w, x \rangle, \langle w', x' \rangle \rangle$ is compatible with what Ralph imagines entails that $\langle w, x \rangle$ is compatible with what Ralph believes. This means that, on our account of *de re* imagining, to say that Ralph imagines that Ortcutt is F is to say something about Ralph's centered belief worlds, namely that there is an acquaintance relation R such that in every one of Ralph's centered belief worlds there is a unique individual to which the center bears R . The content of a *de re* imagining is inextricably connected to the content of one's *de re* beliefs.¹⁰

2.5 *De se* imagining again

At the end of Chapter 1, I noted that in order to solve the problem of the impersonal-*de se*, we would likely need something like two different centers, one representing the *counterfactual-self*, one representing the *belief-self*. It seems that the present proposal furnishes us with exactly what we need. Recall that the counterfactual-self is (to put things intuitively, if imprecisely) who one is imagining being from the inside; the belief-self is the

¹⁰The proposal developed in this section may worry some actualists, for we seem to be re-identifying purely possible individuals across possible worlds. Suppose that $\langle w, x \rangle$ is one of Ralph's centered belief worlds, and that y is the unique thing in w to which x bears R . Now in order to characterize Ralph's imagination alternatives, we make claims about what this same y is like in the centered worlds $\langle w', x' \rangle$ compatible with what Ralph imagines relative to $\langle w, x \rangle$. And of course some of these $\langle w', x' \rangle$ may be such that $w' \neq w$. And note also that y may not be identical to any actual individual.

The problem is that while many actualists don't mind talking about 'non-actual individuals' when those individuals are confined to a single world, they hesitate to re-identify a *particular* non-actual individual across possible worlds (McMichael, 1983). This is a problem, but it is worth noting that the problem isn't one actualists could otherwise avoid, since the problem already arises in connection with sentences like *Kripke could have had a son who was a musician but could have been a plumber*. If a solution to that problem is found, one would hope that that solution could be extended to solve the problem generated by our proposal. Alternatively, we might be able to avoid this problem entirely simply by adopting a counterpart-theoretic approach to the claims in question: it need not be y herself who reappears in w' , but merely a metaphysical counterpart of y , i.e. something that has all of y 's (qualitative) essential properties.

individual in one's imaginary scenario who represents the individual one actually takes oneself to be. Given that we are now representing imaginative possibilities as pairs of centered worlds, it is natural to identify the belief-self with the first center in a pair of centered worlds, and the counterfactual-self with the second center in the pair. The belief-self is imported into the the centered imagination worlds from the centered belief worlds.

The problem of the impersonal-*de se* can be thought of as the problem of characterizing Lingens's imagining in Situation 3. In that situation, amnesiac Lingens has an imagining he reports using (5a):

- (5a) I imagined that I was Brigitte Bardot and that I kissed me.

Take one of Lingens's centered belief worlds, $\langle w, x \rangle$. What must a centered world $\langle w', x' \rangle$ be like in order to be compatible with what Lingens imagines relative to $\langle w, x \rangle$? The obvious answer is that x' (the counterfactual-self) must be Brigitte Bardot in w' , and x' must be kissing x (the belief-self) in w' .¹¹ In other words, the content of the imagining is:

$$\{ \langle \langle w, x \rangle, \langle w', x' \rangle \rangle : x' \text{ is Brigitte Bardot in } w' \text{ and } x' \text{ is kissing } x \text{ in } w' \}$$

And this is a perfectly coherent content. A similar treatment of Situations 1 and 4 can be given. So the present proposal solves the problem of the impersonal-*de se*. Thus, we kill two birds with one stone: our two-dimensional approach solves both the problem of counterfactual attitudes and the problem of the impersonal-*de se*.

Implicit in the above discussion is a characterization of the three types of imagining we discussed in Chapter 1: inside, outside, and the impersonal-*de se*. Let us make this characterization explicit.

In formulating an account of these different kinds of imagining, we again face a choice about whether to characterize the *de se-de dicto* distinction in terms of boring vs. interesting centered propositions or in terms of centered vs. possible worlds propositions (see §1.2.1). In this case, it is simpler to do the former; now that we have two-dimensional intensions, if we did the latter, we would end up with four different kinds of functions (functions from centered worlds to centered propositions, functions from

¹¹Actually, we should first find Bardot's representative in $\langle w, x \rangle$ and then import her (the representative) into $\langle w', x' \rangle$. For simplicity, I ignore this complication.

centered worlds to possible worlds propositions, functions from possible worlds to centered propositions, and functions from possible worlds to possible worlds propositions). To avoid this, we will instead take the contents of all imaginings to be two-dimensional centered intensions, and define two different boring-interesting distinctions for these intensions.

On the new view, we have essentially two ways in which an imagining can be *de se*: it can be about the *me* of my beliefs, or about the *me* of my imaginings. (As the Lakoff case shows, a single imagining can be *de se* in both ways.) To capture this, we need to define two different kinds of *de se* contents, which means we have two different ways for a two-dimensional centered intension to be *interesting*: it can *imagination-interesting* (*i-interesting*), or it can be *belief-interesting* (*b-interesting*). Let's start with the former:

i-boring vs. i-interesting:

- A two-dimensional centered intension p is *i-boring* iff for all w, w' , and x in w and x', y' in w' : $\langle\langle w, x \rangle, \langle w', x' \rangle\rangle \in p$ iff $\langle\langle w, x \rangle, \langle w', y' \rangle\rangle \in p$.
- A two-dimensional centered intension is *i-interesting* iff it is not *i-boring*.

Using these terms, we can characterize the inside-outside distinction as follows:

2D Uncentered Imagination

The content of an imagining from the outside an *i-boring* two-dimensional centered intension.

For all imaginings I , I is an imagining from the outside just in case the content of I is an *i-boring* two-dimensional centered intension.

2D Centered Imagination

The content of an imagining from the inside an *i-interesting* two-dimensional centered intension.

For all imaginings I , I is an imagining from the inside just in case the content of I is an *i-interesting* two-dimensional centered intension.

The content of an outside imagining – an *i-boring* two-dimensional centered intension – is a function from centered worlds to boring centered

propositions, (which is more or less equivalent to a function from centered worlds to possible worlds propositions). Given one of the subject's centered belief worlds, the function gives back a boring centered proposition, which represents the fact that the imagining doesn't engage the counterfactual-self. The imagining is not 'from the perspective' of anyone in the imagined scenario, though it might be about the belief-self.

In order to characterize impersonal-*de se* imagining, we need the notion of a *b*-interesting (*b*-interesting) two-dimensional centered intension:

b-boring vs. *b*-interesting:

- A two-dimensional centered intension p is *b*-boring iff for all w, w' , and x and y in w, x' in w' : $\langle\langle w, x \rangle, \langle w', x' \rangle\rangle \in p$ iff $\langle\langle w, y \rangle, \langle w', x' \rangle\rangle \in p$.
- A two-dimensional centered intension is *b*-interesting iff it is not *b*-boring.

A *b*-boring two-dimensional centered intension is a function p from a centered world $\langle w, x \rangle$ to a centered proposition, where p isn't sensitive to the value of x : it returns the same centered proposition no matter what value is assigned to x . The x -coordinate is idle, and so the function is essentially equivalent to a function from possible worlds to centered propositions. This reflects the fact that if a *b*-boring two-dimensional centered intension were the content of an imagining, the imagining wouldn't be about the agent's belief-self. Using this notion we can characterize impersonal-*de se* imaginings:

2D Impersonal-*De Se* Imagination

The content of an impersonal-*de se* imagining is a *b*-interesting two-dimensional centered intension.

For all imaginings I , I is an impersonal-*de se* imagining just in case the content of I is a *b*-interesting two-dimensional centered intension.

Note that on this taxonomy, any imagining is either from the inside or from the outside, but never both. But an imagining can be from the inside and impersonal-*de se*, as in the Bardot case, as well as in more mundane cases when one imagines being oneself (i.e. when the counterfactual-self and the belief-self coincide). An imagining can also be from the outside

2. Imagination, Descriptivism, and the *De Re*

and impersonal-*de se*, as when, for example, amnesiac Lingens imagines from the outside that he himself is skiing down a steep hill (see Situation 1 in §1.3).

According to our account, the contents of counterfactual attitudes will be two-dimensional centered intensions, while the contents of beliefs (and belief-like attitudes such as knowledge) could be represented as a centered propositions, as they are on Lewis's view of *de re* belief. However, if one desired that the objects of all attitudes be things of the same kind, we could instead take the objects of belief to be two-dimensional centered intensions; one of the 'dimensions' will simply be idle.¹²

Whether or not one accepts the idea that the content of an imagining should be characterized as a two-dimensional centered intension, I think

¹²The idea of characterizing the content of an attitude as a two-dimensional intension may also be needed by Chalmers/Jackson-style two-dimensionalists, since they seem to face their own problem of counterfactual attitudes. To see this, suppose the following things:

- I am ignorant of the fact that Hesperus is Phosphorus.
- My 1-intension for *Hesperus* is: λw . the celestial body that appears in the morning in w .
- My 1-intension for *Phosphorus* is: λw . the celestial body that appears in the evening in w .
- I believe (truly) that it isn't an essential property of Hesperus that it appears in the morning.
- I believe (truly) that it isn't an essential property of Phosphorus that it appears in the evening.

If that's what my belief state is like, then it would seem that I can imagine:

that Hesperus is not Phosphorus, and that Hesperus does not appear in the morning, and that Phosphorus does not appear in the evening.

But this has an empty 1-intension, since there is no world w in which the celestial body that appears in the morning in w does not appear in the morning in w . And it also has an empty 2-intension, since there is no world w in which the thing that is actually Phosphorus (i.e. Venus) is distinct from the thing that is actually Hesperus (Venus).

The way out is to take the content of the imagining to be the relevant two-dimensional intension:

$\{(w, w') : \text{the celestial body that appears in the morning in } w \text{ is not the celestial body that appears in evening in } w, \text{ and the celestial body that appears in the morning in } w \text{ does not appear in the morning in } w', \text{ and the celestial body that appears in the evening in } w \text{ does not appear in the evening in } w'\}$

(Cf. Chalmers (2002b, 163) on the two-dimensional content of sentences like: "If water is XYZ, then water could not be H_2O .)

that if one is a descriptivist about the *de re*, then one should accept the basic intuition driving our theory: that counterfactual attitudes ought to be understood in relation to the ‘base’ attitude of belief. There may be other ways of cashing out this basic intuition; if there are, it would be interesting to compare them to the present proposal. But I leave this as a matter for future inquiry.

2.6 Attitude ascription

2.6.1 The Lewisian account

While Lewis’s account of *de se* attitudes is popular in the philosophical literature, his account of the *de re* is less so, not least because descriptivism about the *de re* has fallen out of favor in recent years (a point I will discuss a bit more at the beginning of the next chapter). But Lewis’s account of the *de re* has inspired a certain approach to the semantics of *de re* ascription that has some claim to being the dominant approach to *de re* ascription among formal semanticists. There are a number of different accounts in this vein, but all are essentially variants of Cresswell and von Stechow (1982).¹³ The particular formulation I’ll rely on is based on one discussed in Anand (2006, Ch.1). The basic semantic framework used here is the same as the one introduced in §1.2; we also retain the account of *de se* ascription discussed in that section.

Consider sentence (9):

- (9) Ralph believes that Ortcutt is a spy.¹⁴

On Lewis’s account, this sentence is true at a world *w* just in case there is an acquaintance relation *R* that Ralph bears uniquely to Ortcutt, and is such that that Ralph believes (*de se*) that the one to whom he bears *R*

¹³See, for example, von Stechow (1982), Heim (1994a), Abusch (1997), Anand (2006), and Maier (2006).

¹⁴In my discussion, I assume that attitude reports in which proper names, indexicals, and pronouns occur within the scope of the attitude verb report *de re* attitudes. This is not uncontroversial, since our practices of attitude ascription may allow us to use an attitude report of that type even when the agent is not acquainted with the *res* in the sense discussed at the outset of this chapter. One could respond to this in one of two ways: one could relax the acquaintance requirements on *de re* attitudes considerably; or one could deny that every attitude report in which proper names, indexicals, and pronouns occur within the scope of the attitude verb reports an attitude that is genuinely *de re*. See Sosa (1970) Chisholm (1976), Burge (1977, 346), Chisholm (1981, Ch.9), Pryor (2004), and Anand (2006, 20) for discussion.

uniquely is a spy. How do we predict this truth condition for (9)? What would the semantic value of *believes* have to be like? What would the LF of (9) have to be like?

The theory we adopt assumes that the LF of (9) is given by (10):

- (10) Ralph [[believes Ortcutt] [$\lambda_1 \lambda_2 t_2$ is a spy]]

Let me make a few comments about this structure. First, the account hypothesizes that *Ortcutt* moves out of the lower clause and into an argument position of the attitude verb. When *Ortcutt* moves, it leaves behind a trace, here ' t_2 ', which is bound by the variable binder ' λ_2 ', so that the third argument of the attitude verb is a 'centered property' (a function from individuals to centered propositions) rather than a centered proposition. The variable binder ' λ_1 ' does not actually bind anything in this particular structure; it is present to ensure that the third argument is a centered property.

The lexical entry for *believes* looks like this:

$\llbracket \text{believes} \rrbracket^{c,i,g} = \lambda res_e. \lambda p_{\langle s, \langle e, et \rangle \rangle}. \lambda att_e. \text{there is an acquaintance relation } R \text{ such that } R(w_i)(att) = res, \text{ and all the centered worlds } \langle w, x \rangle \text{ compatible with what } att \text{ believes at } w_i \text{ are such that } p(w)(x)(R(w)(x)) = 1.$ ¹⁵

Here we take relations of acquaintance to be two-place functions that take a possible world and an individual and deliver an individual; $R(w)(x)$ = the unique individual y to whom x bears R in w . With the LF in (10) and the above semantic value for *believes*, we end up with the following truth conditions for (9):

$\llbracket \text{Ralph believes Ortcutt } \lambda_1 \lambda_2 t_2 \text{ is a spy} \rrbracket^{c,i,g} = 1 \text{ iff}$

There is an acquaintance relation R such that $R(w_i)(\text{Ralph}) = \text{Ortcutt}$, and all the centered worlds $\langle w, x \rangle$ compatible with what Ralph believes at w_i are such that $\lambda w'. \llbracket \lambda_1 \lambda_2 t_2 \text{ is a spy} \rrbracket^{c,i,g}(w)(x)(R(w)(x)) = 1$ iff

There is an acquaintance relation R such that $R(w_i)(\text{Ralph}) = \text{Ortcutt}$, and all the centered worlds $\langle w, x \rangle$ compatible with what Ralph believes at w_i are such that $R(w)(x)$ is a spy in w' .

¹⁵This entry only deals with the 'singly *de re*' case, i.e. the case in which there is only one *res* expression in the complement clause. I will only be discussing this case. For a suggestion on how to generalize the account, see Cresswell and von Stechow (1982).

And this is precisely the result we want.

One interesting feature of this approach is that it combines a ‘direct reference’ semantics for names, pronouns, and indexicals with descriptivism about attitude contents. Though this account of *de re* ascription is often discussed in the formal semantics literature (as I noted earlier), the possibility of combining this type of account of referring expressions with a descriptivist account of attitudes seems to have been largely overlooked in the philosophical literature dealing with these topics.¹⁶

It should be clear that the problem of counterfactual attitudes extends to this theory of *de re* ascription. The problem is not with the account of *believes*, but with extending the theory to cover counterfactual attitude verbs like *wishes*, *imagines*, and *dreams*. Consider sentence (11):

- (11) Ralph imagined that he had never been acquainted with Ortcutt.

I take it that this sentence is true in the situation we described in §2.3, in which Ralph imagines a scenario in which Ortcutt and he never cross paths. But this semantic theory incorrectly predicts that (11) is false in that situation. Indeed, this theory predicts that there is *no* world at which (11) is true (it is necessarily false). For consider any world w :

$\llbracket \text{Ralph imagined Ortcutt } \lambda_1 \lambda_2 \text{ he}_1 \text{ had never been acquainted with } t_2 \rrbracket^{c,w,g} = 1$ iff

There is an acquaintance relation R such that $R(w)(\text{Ralph}) = \text{Ortcutt}$, and all the centered worlds $\langle w', x' \rangle$ compatible with what Ralph imagines at w are such that $\lambda w''. \llbracket \lambda_1 \lambda_2 \text{ he}_1 \text{ had never been acquainted with } t_2 \rrbracket^{c,w'',g}(w')(x')$

$(R(w')(x')) = 1$ iff

There is an acquaintance relation R such that $R(w)(\text{Ralph}) = \text{Ortcutt}$, and all the centered worlds $\langle w', x' \rangle$ compatible with what Ralph imagines at w are such that x' is never acquainted with $R(w')(x')$ in w' .

The problem with this is that R is acquaintance relation, and $R(w')(x')$ is the individual to whom x' bears R in w' . Thus, the sentence is true only if there is an acquaintance relation R such that all the $\langle w', x' \rangle$ compatible with Ralph imagines in w are such that x' is not acquainted in w with

¹⁶Though see Shier (1996) and Bach (1997).

the individual to whom x' bears R in w . But no acquaintance relation meets that condition, and so the sentence is false at w . But since w was an arbitrary world, the same holds for every world, and so (11) comes out necessarily false on this account. But of course (11) is true at some worlds: we described one such world earlier. So this account of *imagines* should be abandoned. Similar counterexamples could be given for the corresponding accounts of *wishes* and *dreams*.

2.6.2 A two-dimensional refinement

In light of our revision to (or perhaps extension of) Lewis's theory, we can see where a solution lies. There might be a number of ways to change the semantics so that it avoids the problem of counterfactual attitudes but the least invasive, most conservative route is simply to change the lexical entry of *imagines* as follows:

$\llbracket \text{imagines} \rrbracket^{c,i,g} = \lambda res_e. \lambda p_{\langle s, \langle e, et \rangle \rangle}. \lambda att_e. \text{there is an acquaintance relation } R \text{ such that } R(w_i)(att) = res, \text{ and every } \langle \langle w, x \rangle, \langle w', x' \rangle \rangle \text{ compatible with what } att \text{ imagines at } w_i \text{ is such that there is a } y \text{ such that } R(w)(x) = y, \text{ and } p(w')(x')(y) = 1.$

If we adopt this revision, we needn't even revise the semantic type of a counterfactual attitude verb. On this theory, while the contents of imaginings are two-dimensional centered intensions, the complement of *imagines* is still a function from individuals to centered propositions. Thus, a nice feature of our account of imagining is that it requires only a minimal change to our preferred account of the semantics of attitude ascriptions.

Given the above lexical entry, if the LF of (11) is:

Ralph $\llbracket \text{[imagined Orcutt]} [\lambda_1 \lambda_2 \text{ he}_1 \text{ had never been acquainted with } t_2] \rrbracket$

then we can compute the truth conditions of (11) as follows:

$\llbracket \text{Ralph imagined Orcutt } \lambda_1 \lambda_2 \text{ he}_1 \text{ had never been acquainted with } t_2 \rrbracket^{c,w,g} = 1 \text{ iff}$

There is an acquaintance relation R such that $R(w)(\text{Ralph}) = \text{Orcutt}$, and every $\langle \langle w', x' \rangle, \langle w'', x'' \rangle \rangle$ compatible with what Ralph imagines at w is such that there is a y' such that $R(w')(x') = y'$, and $\lambda w'''. \llbracket \lambda_1 \lambda_2 \text{ he}_1 \text{ had never been acquainted with } t_2 \rrbracket^{c,w''',g}(w'')(x'')(y') = 1$ iff

There is an acquaintance relation R such that $R(w)(\text{Ralph}) = \text{Ortcutt}$, and every $\langle\langle w', x' \rangle, \langle w'', x'' \rangle \rangle$ compatible with what Ralph imagines at w is such that there is a y' such that $R(w')(x') = y'$, and x'' is not acquainted with y' in w'' .

Since there are worlds that meet this condition, the sentence is not necessarily false on this semantics.

Finally, let me say something about the impersonal-*de se*. Recall Situation 1 from §1.3. In that Situation, amnesiac Lingens imagines, in the impersonal-*de se* way, that he himself is skiing down a steep hill, i.e. he imagines this from the outside. How would we report this imagining? We'd report it by saying something like:

- (12) Lingens imagined that he was skiing down a steep hill.

Given what we've said about the semantics thus far, (12) is ambiguous between two structures, one in which *he* is bound, and one in which it's free (recall §1.2). When *he* is bound, the inside reading is generated. When *he* is free, our new account of *de re* ascription tells us that it moves into an argument position closer to the verb. Here are the two structures:

- Lingens imagined $\lambda_1 \text{ he}_1$ was skiing down a steep hill. (bound, ‘inside reading’)
- Lingens imagined $\text{he}_1 \lambda_2 t_2$ was skiing down a steep hill. (free, ‘*de re* reading’)

Now the bound structure of (12) is not true in Situation 1 because Lingens isn't imagining skiing from the inside. So if that sentence has true reading in Situation 1, it must be the free structure that is true, which gives us the *de re* reading. How then can (12) be true in that Situation, given that Lingens's imagining is impersonal-*de se*? It can be true in that Situation because the impersonal-*de se* can be understood as a special case of the *de re*. The trick is to allow the relation of identity to count as a relation of acquaintance; after all, as Lewis (1979, 156) writes, “identity is a relation of acquaintance par excellence.”

According to our account of the impersonal-*de se*, Lingens imagines, in the impersonal-*de se* way, that he is skiing down a steep hill iff all the $\langle\langle w, x \rangle, \langle w', x' \rangle \rangle$ compatible with what he imagines are such that x is skiing down a steep hill in w' . But then Lingens imagines, in the impersonal-*de se* way, that he is skiing down a steep hill iff:

- Lingens bears the relation of identity to Lingens; and
- all the $\langle\langle w, x \rangle, \langle w', x' \rangle \rangle$ compatible with what he imagines are such that there is a unique y in w such that x bears the relation of identity to y , and y is skiing down a steep hill in w' .

But that means that if Lingens imagines, in the impersonal-*de se* way that he is skiing down a steep hill, then Lingens imagines *de re* that he is skiing down a steep hill. Impersonal-*de se* imagining is *de re* imagining where the acquaintance relation is identity.

So (12) can be true in Situation 1, because (12) has a reading on which it is true iff Lingens imagines *de re* that he is skiing down the hill. Since impersonal-*de se* imagining is a species of *de re* imagining, then if Lingens impersonally-*de se* imagines skiing down a steep hill – as he does in Situation 1 – sentence (12) will be true on its *de re* reading.¹⁷

2.6.3 Acquaintance relations: existential quantification vs. contextually supplied?

On the proposed treatment of *de re* ascriptions, those ascriptions involve existential quantification over relations of acquaintance. Some theorists think that *de re* ascriptions are context sensitive in a certain way (Abusch, 1997, n. 9): on their view, (9) for example will be true in a context in which the ‘trenchcoat’ acquaintance relation is salient, and false in one in which the ‘mayor’ acquaintance relation is salient.¹⁸

(9) Ralph believes that Ortcutt is a spy.

But on the view we’ve taken, that sentence will be true in any context as long as there is *some* acquaintance relation R such that Ralph believes that the one to whom he bears R uniquely is a spy.

I don’t know which of these views is correct (see Anand (2006, Ch.1) for some discussion). But the ‘contextualist’ view can be obtained rather easily from our story by adding that the domain of quantification can be

¹⁷Note that on our approach, the impersonal-*de se* is not linguistically marked, in the sense that there are no structures that unambiguously express impersonal-*de se* imagination reports.

¹⁸This dispute parallels one in the philosophy of language literature between *neo-Russellians* like Salmon (1986), who think belief ascriptions quantify over modes of presentation, and *hidden-indexical theorists* like Schiffer (1977) and Crimmins (1992), who think modes of presentation are contextually supplied.

contextually restricted. The change to the semantics is minor: one need only add that attitude verbs take as another argument a set of acquaintance relations. This could be represented in the structure as a variable over such sets, and the context could be relied on to determine which set gets assigned to that variable.

Allowing the variable to range over sets of acquaintance relations rather than over the relations themselves gives us more flexibility. The context can, but need not, determine *one particular* acquaintance relation. If the participants of the conversation do have in mind a particular acquaintance relation R , the context can assign the singleton set $\{R\}$ to the variable. If they have no particular acquaintance relations in mind at all, the context can assign the set of all acquaintance relations to the variable. If, however, their intentions and information rule out some, but not all, acquaintance relations, the set assigned to the variable will contain all and only those relations not ruled out by their intentions and information. The resulting ascription will say that one of those relations R is such that the attitude holder bears R to the *res*, and believes that the unique thing to which he bears R is such-and-such.

Chapter 3

Imagination, Sequenced Worlds, and the *De Re*

Descriptivism about *de re* attitudes is controversial. Many philosophers think an agent can have *de re* thoughts about an object *x* even if she doesn't possess a descriptive concept that uniquely identifies *x*. Burge gives this example:

On seeing a man coming from a distance in a swirling fog, we may plausibly be said to believe of him that he is wearing a red cap. But we do not see the man well enough to describe or image him in such way as to individuate him fully. (Burge, 1977, 351 - 352)

If there is one man, I might think of him under the descriptive concept *the man I see in the red cap*. But what if there are multiple red-capped men? I focus on one, but know of no property that picks him out uniquely. There is a failure of uniqueness: I have some information about the *res*, but that information may not uniquely fit one particular object that I am acquainted with. Maybe if we scrutinize the case further, we will be able to come up with a descriptive concept which (a) picks out the man in question uniquely, and (b) is a plausible candidate for the descriptive concept I use in thinking about the man. But what guarantee is there that every case will be like this?

Another type of case that poses a *prima facie* problem for descriptivism: Ralph sees someone at a bar with a drink in hand. He thinks to himself, *the man with the gin and tonic has a colorful hat*. In fact, it is a woman drinking sparkling water (cf. Donnellan, 1966). Despite the fact that that woman does not satisfy Ralph's descriptive concept, it seems plausible to say that Ralph has a *de re* belief about her, a judgment that appears to conflict with descriptivism. Again, maybe Ralph possesses a 'fallback'

descriptive concept which the woman satisfies; but, again, how can we be sure an agent will *always* have a fallback descriptive concept which the *res* uniquely satisfies? Descriptivism can begin to look like a bold conjecture, one we have little reason to accept.

These considerations are not decisive. There is a longer debate to be had, more examples to haggle over. Maybe descriptivism will emerge unscathed.¹ In the last chapter, we went along with Lewis's descriptivism about the *de re*, showing how we could solve both the problem of counterfactual attitudes and the problem of the impersonal-*de se* within a descriptivist framework. But given the controversy surrounding descriptivism, it would be interesting to see what one should say about the *de re* if one accepts Lewis's view of the *de se* but rejects descriptivism.

This chapter shows how Lewis's account of the *de se* can be generalized to give a non-descriptivist account of the *de re*. On Lewis's account of the *de se*, we *stipulate* that the center represents the agent in any one of the agent's centered belief worlds. I show how to extend this 'method of stipulation' to *de re* attitudes. The resulting account solves the problem of counterfactual attitudes and the problem of the impersonal-*de se*, and yields a non-descriptivist account of *de re* attitude ascriptions.

3.1 Identification and stipulation

Let me rely for the moment on our pre-theoretic understanding of the terms 'possible situation' and 'doxastic alternative' (a doxastic alternative is just a possible situation compatible with what someone believes). Thus we leave it open for the time being whether possible situations ought to be understood as possible worlds, or as centered worlds, or as something else. One of the main points of this section concerns how non-descriptivists about the *de re* should understand the notion of a doxastic alternative.

When Ralph has a *de re* belief about Ortcutt in virtue of bearing *R* to him, each of Ralph's doxastic alternatives *s* contains someone who represents Ortcutt. The question of '*de re* identification' is this: how do we identify the individual in *s* who represents Ortcutt.² Lewis's descriptivist account of the *de re* offers an answer to this question: *s* is a centered world, and the individual *y* in *s* who represents Ortcutt is the one to whom the

¹For a recent defense of (what I am calling) descriptivism against similar objections, see Chalmers (2002a,b).

²Hintikka (1969, 1975) first raised this sort of question – the question of 'cross-identification' – in the context of possible worlds semantics.

3. Imagination, Sequenced Worlds, and the *De Re*

center bears *R* uniquely. If descriptivism is rejected then we must find a different answer to the question of *de re* identification.

The Naive Theory also offers an answer to this question: the one who represents Ortcutt in one of Ralph's doxastic alternatives is Ortcutt himself. But we've already seen reasons for rejecting the Naive Theory: it cannot handle the puzzles about identity, and it runs into a problem concerning the essential properties of the *res* (see §2.1).³ If both descriptivism and the Naive Theory are rejected, how are we to answer the question of *de re* identification? I think an answer to that question can be found if we look more closely at how Lewis answers the corresponding question of '*de se* identification'. When Ralph has a *de se* belief about himself, each of his doxastic alternatives *s* contains an individual who represents Ralph. The question of *de se* identification is: how do we identify the individual in *s* who represents Ralph?

Lewis's answer goes something like this: *s* is a centered world $\langle w, x \rangle$, and *x* is its center. It is, I think, an implicit *stipulation* of Lewis's account that the center of a centered world represents the attitude holder in that centered world.⁴ The centered worlds strategy seems to consist of two parts: (i) individuating possibilities more finely: doxastic possibilities are now individuals-at-possible-worlds, rather than merely possible worlds; and (ii) an implicit stipulation that, when we consider a centered world as a candidate doxastic alternative for some agent *x*, the center of that centered world represents *x*.

Let me explain (ii): any centered world $\langle w, x \rangle$ is a candidate doxastic alternative for me, in the sense that it is the sort of thing we can assess for compatibility with my belief state. When we are assessing whether $\langle w, x \rangle$ is compatible with my belief state, we check to see whether *x* in *w* has all the properties I attribute to myself. If it does, then it's compatible with what I believe; if it doesn't, then it's not. That's the sense in which the center of a centered world represents me in that centered world, when we're assessing that centered world for compatibility with my belief state. (When we assess a centered world for compatibility with *Ralph's* belief state, we

³What about a counterpart-theoretic variation of the Naive Theory, one on which the one who represents Ortcutt in a doxastic alternative is a *metaphysical counterpart* of Ortcutt? This avoids one of the problems that faced the Naive Theory, since Ortcutt could have more than one counterpart in one of Ralph's doxastic alternatives, but it won't help with the problem of essential properties, since Ortcutt's metaphysical counterparts share all his essential properties (cf. Lewis, 1981).

⁴See the discussion of 'finding the subject by stipulation' in Lewis (1983a). (Talk of stipulation is more prominent in Lewis (1983a) than in Lewis (1979).)

understand the center as representing *Ralph* in that centered world. The generalization is that when we assess a centered world for compatibility with an agent's belief state, the center represents the agent.)

But note that we can quite easily extend this same strategy to my *de re* beliefs about others. Suppose for simplicity that I only have beliefs about one other individual, call him "Sam". In that case, in every possible situation s compatible with what I believe, there will be an individual who represents me, and an individual who represents Sam. (Assuming I believe that I am not Sam, these individuals will be distinct.) How should we represent these possible situations? Taking our cue from the centered worlds strategy, we could (i) individuate possibilities still more finely: a doxastic possibility will now be a pair of a possible world w and a *pair* of individuals who inhabit w , rather than merely a centered world; and (ii) stipulate that, when we consider an object like $\langle w, \langle x, y \rangle \rangle$ as a candidate doxastic alternative for me, x represents me in $\langle w, \langle x, y \rangle \rangle$, and y represents Sam in $\langle w, \langle x, y \rangle \rangle$. Point (i) tells us to represent my candidate doxastic alternatives using objects of a certain sort. Point (ii) tells us how to understand what these objects are being used to represent.

I'll say more momentarily about how to generalize point (ii), but let me first try to give you an intuitive feel for the underlying idea. Call a pair consisting of a possible world and a pair of individuals in that world a *sequenced world* (a pair is a sequence of length two).⁵ Any sequenced world is a candidate for being one of my doxastic alternatives: it is the sort of thing we can assess for compatibility with my belief state. What does it take for a sequenced world to be compatible with what I believe? In our toy example in which I only have beliefs about myself and Sam, a sequenced world $\langle w, \langle x, y \rangle \rangle$ will be compatible with what I believe iff:

- x in w has all the properties that I attribute to myself;
- y in w has all the properties I attribute to Sam; and
- y and x in w stand in all the relations that I believe that Sam and I stand in.

(Remember we are stipulating that x represents me, and that y represents Sam.) So, for example, if I believe:

- that I am small;

⁵I owe the term *sequenced world* to Seth Yalcin.

3. Imagination, Sequenced Worlds, and the *De Re*

- that Sam is big; and
- that Sam and I are friends

then all the sequenced worlds $\langle w, \langle x, y \rangle \rangle$ compatible with what I believe will be such that:

- x is small in w ;
- y is big in w ; and
- y and x are friends in w

The ‘representing individual’ in a doxastic alternative always has all the properties that the agent attributes to the corresponding ‘represented individual’.

Another way of explicating our proposal uses the metaphor of *self-location* (Lewis, 1979). On Lewis’s account of the *de se*, believing something is locating oneself somewhere in the space of possible individuals. If I believe that I’m a philosopher, I locate myself in the set of possible individuals who are philosophers: I believe that I am one of the people in that set. Our suggestion is that, in addition to self-location, there is something that deserves the name ‘*other-location*’. When I believe something about Sam and myself, I am locating *both* of us in the space of possible individuals. Better: I am locating the *pair* consisting of me and Sam in the space of *pairs* of possible individuals (where the members of the pairs are world-mates). I am *self-and-other-locating*, or *self-and-Sam-locating*. If I believe that Sam and I are friends, then I locate $\langle \text{Dilip}, \text{Sam} \rangle$ in the set of pairs of possible individuals who are friends. If I believe that I am hungry and Sam is thirsty, I locate $\langle \text{Dilip}, \text{Sam} \rangle$ in the set of pairs of possible individuals whose first member is hungry and whose second member is thirsty. I believe that $\langle \text{Dilip}, \text{Sam} \rangle$ is in that set.

The language of self- and other-location can be translated into the language of doxastic compatibility. A centered world is a way of representing a possible individual; so Lewis:

...centered worlds are not just worlds, just as stuffed cabbage is not just cabbage; centered worlds amount to presentations of possible individuals, so a proposal to use centered worlds

differs little from my proposal to use the individuals themselves.
 (Lewis, 1983a, n. 18)⁶

If centered worlds are ‘presentations of possible individuals’, then when I self-locate, we can say that what I’m doing is locating the centered world (actual world, Dilip) in a set of centered worlds. A centered world $\langle w, x \rangle$ is compatible with what I believe in the actual world just in case I locate (actual world, Dilip) in a set containing $\langle w, x \rangle$: my beliefs don’t rule out the possibility that I am x in w , and so $\langle w, x \rangle$ represents a way I might be.

Similarly, a sequenced world is way of representing a pair of possible individuals. So, on our proposal, when I believe something, I locate (actual world, \langle Dilip, Sam \rangle) in a set of sequenced worlds. A sequenced world $\langle w, \langle x, y \rangle \rangle$ is compatible with what I believe in the actual world just in case I locate (actual world, \langle Dilip, Sam \rangle) in a set containing $\langle w, \langle x, y \rangle \rangle$: my beliefs don’t rule out the possibility that I am x in w and Sam is y in w , and so $\langle w, \langle x, y \rangle \rangle$ represents a way me and Sam might be.⁷

Talk of pair-location encourages us to make our stipulation about who x and y represent in $\langle w, \langle x, y \rangle \rangle$ more explicit: it is me and Sam who I am locating. Thus, in our simple example, instead of talking about what it is for a sequenced world $\langle w \langle x, y \rangle \rangle$ to be compatible with what I believe *simpliciter*, we should talk about what it is for a sequenced world to be compatible with what I believe in the actual world about (Dilip, Sam). When we assess whether $\langle w, \langle x, y \rangle \rangle$ is compatible with what I believe in the actual world about (Dilip, Sam), we understand x as representing Dilip and y as representing Sam (the n th member of $\langle x, y \rangle$ represents the n th member of (Dilip, Sam)). This will be more important to keep in mind when we move to sequences of greater length.

The *content* of my belief that Sam is big can be represented by the following object:

$$\langle \langle \text{actual world}, \langle \text{Dilip, Sam} \rangle \rangle, \{ \langle w, \langle x, y \rangle \rangle : y \text{ is big in } w \} \rangle$$

⁶In order to use individuals instead of centered worlds (world-individual pairs), one needs counterpart theory’s metaphysic of world-bound individuals (for then an individual determines a possible world). In a system with strict identity across worlds, individuals alone will not suffice. For suppose I believe that I am resilient. Now consider the question: is the *individual* Hillary Clinton compatible with what I believe? In a system of strict identity, there’s no answering this question, since Hillary is resilient in some worlds, not so in others, i.e. Hillary-at- w is, Hillary-at- w' isn’t. Thus the need for centered worlds (individuals-at-worlds), rather than mere individuals.

⁷Forgive the ‘me and Sam’ locution, but ‘Sam and I’ would be misleading in this context and ‘I and Sam’ sounds awful.

3. Imagination, Sequenced Worlds, and the *De Re*

Here, a content is understood to be a pair that consists of a sequenced world and a set of sequenced worlds.⁸ Note that the first member of the pair is *maximal* in the sense that it is a sequenced world whose second member contains all and only those individuals I have beliefs about (namely Dilip and Sam). If p is a set of sequenced worlds, to say that $\langle\langle w, \langle \text{Dilip, Sam} \rangle \rangle, p \rangle$ is the content of my belief is to say that all the sequenced worlds compatible with what I believe in w about $\langle \text{Dilip, Sam} \rangle$ are contained in p (I locate $\langle w, \langle \text{Dilip, Sam} \rangle \rangle$ in p). If I believe $\langle\langle w, \langle \text{Dilip, Sam} \rangle \rangle, p \rangle$, then my belief is *true* iff $\langle w, \langle \text{Dilip, Sam} \rangle \rangle \in p$.⁹

Continuing with our toy example, we can offer this preliminary account of *de re* belief:

Dilip believes (*de re*) in w that Sam is F iff

- i. there is an acquaintance relation R such that Dilip bears R to Sam in w ; and
- ii. all the sequenced worlds $\langle w', \langle x', y' \rangle \rangle$ compatible with what Dilip believes in w about $\langle \text{Dilip, Sam} \rangle$ are such that y' is F in w' .

Let us call $\langle \text{Dilip, Sam} \rangle$ a *res sequence* for me in w : it is a sequence that contains all and only the individuals I have beliefs about in w , a sequence of all and only the individuals I am locating in the space of sequences of possible individuals.

⁸Since sets of sequenced worlds are essentially relations, contents are a kind of structured proposition, not too distant from the sort of contents favored by many philosophers of language (see King (2006) for an overview).

⁹In his unpublished Locke Lectures on self-locating belief, Robert Stalnaker takes the content of a *de se* belief to be a pair consisting of a centered world and a set of centered worlds, where the first member of the pair is the centered world that correctly locates the agent. Our proposal bears affinities to Stalnaker's idea. See also the distinction in Recanati (2007b) between the *explicit content* of a thought and its associated *Austinian proposition*. In our system, what we are calling the content of a belief would be the Austinian proposition; the associated set of sequenced worlds would be the explicit content. An Austinian proposition has a truth value *simpliciter*, i.e. not relative to anything, not even a possible world.

If we adopted Recanati's distinction, we might end up with a non-descriptivist notion of *narrow content*; narrow contents in our system would be sets of sequenced worlds, and they would have truth values relative to a sequenced world. (In order for this to be a genuine notion of a narrow content, we might have to extend this proposal to *properties* as well as individuals.) I hope to explore these ideas in future work. (Thanks to Seth Yalcin for pressing me to think about what the contents of belief should be on this account.)

This account avoids descriptivism because we do not use a descriptive concept to help us find the individual who represents the *res* in a doxastic alternative: we find the *res*'s representative by stipulation. One way to put this is that if we're just given a centered world $\langle w, x \rangle$, and told that it is compatible with what I believe about myself, we cannot necessarily recover a sequenced world $\langle w, \langle x, y \rangle \rangle$ from it, because the individual *y* who represents Sam in one of my doxastic alternatives need not be 'obtainable' as function of *w* and *x*. This reflects the fact that I might not be thinking of Sam under some descriptive concept *the G*, and so my belief state might fail to determine a function from centered worlds to representatives of Sam.

One question that one might have about our proposal is this: what *makes* it the case that *y* in $\langle w, \langle x, y \rangle \rangle$ represents Sam when we assess whether that sequenced world is compatible with what I believe about \langle Dilip, Sam \rangle ? In essence, no substantive answer to this question can be given: we have stipulated that *y* represents Sam in $\langle w, \langle x, y \rangle \rangle$ when we are assessing whether that sequenced world is compatible with what I believe about \langle Dilip, Sam \rangle . Compare the question: what *makes* it the case that the center of a centered world represents Dilip when we are assessing whether that centered world is compatible with what Dilip believes? Here, too, it seems as if no substantive answer can be given.

Assuming the method of stipulation is legitimate in the *de se* case, is it also legitimate in the *de re* case? One might think the two cases are different in virtue of the fact that *de se* and third-personal *de re* thought are fundamentally different in certain respects. It is widely believed that *de se* thought is in some ways distinctive, and that it differs from ordinary third-personal *de re* thought in important ways. After all, isn't that what Perry, Lewis, and others showed when they showed that *de se* thoughts are irreducibly indexical? I do think there is something distinctive about the *de se*, though it can be tricky to say just what this is (it seems to have something to do with the distinctive role of *de se* thought in the explanation of action).¹⁰ But what is not at all clear is *why* the distinctiveness of the *de se* makes the method of stipulation legitimate in the *de se* case, but not in the *de re* case. What, exactly, is the connection between the distinctive nature of *de se* thought and the method of stipulation?

This isn't meant as a criticism of Lewis's treatment of the *de se*. But the best argument for his account is not that it captures the distinctiveness of the *de se*. In my view, the best argument for his account is simply that

¹⁰I think that reading Perry (1977, 1979) is the best way to get a sense of the distinctiveness of the *de se*.

it gives us an elegant representation of *de se* thought, one which, being a generalization of the familiar possible worlds treatment of attitudes, makes it easier to see how to theorize about the *de se* in other domains, like formal semantics (as we have seen) and decision theory (e.g. in thinking about the Sleeping Beauty problem). But these reasons for being interested in Lewis's account of the *de se* do not suggest that that account cannot be fruitfully generalized into a theory of the *de re*.

So we have not, I think, been given a reason to refrain from using the method of stipulation when theorizing about the *de re*. Thus, we press on with the task of developing our sequenced worlds account of *de re* attitudes.

3.2 Identity puzzles and acquaintance relations

Two of the main advantages of Lewis's descriptivist account of the *de re* over the Naive Theory are its ability to treat the familiar puzzles of identity confusion, and its ability to better represent beliefs about essential properties. Here we show that our anti-descriptivist alternative to Lewis's account can also do this work.

Let me start with essential properties. Suppose Ortcutt has DNA sequence d , and has that sequence essentially. Nevertheless, Ralph might not have any opinions about Ortcutt's DNA sequence. Our account can easily represent this because there is no requirement that the individual who represents Ortcutt in one of Ralph's doxastic alternatives be Ortcutt himself or a metaphysical counterpart of Ortcutt. On our account, some of the $\langle w, \langle x, y \rangle \rangle$ compatible with what Ralph believes about $\langle \text{Ralph}, \text{Ortcutt} \rangle$ are such that y has DNA sequence d in w , and some of the $\langle w', \langle x', y' \rangle \rangle$ compatible with what Ralph believes about $\langle \text{Ralph}, \text{Ortcutt} \rangle$ are such that y' doesn't have DNA sequence d in w' .¹¹ (Note that y and y' are allowed to be distinct.) So our account avoids this sort of worry involving essential properties.¹²

Cases of identity confusion – the Ortcutt case, the Lingens case – show

¹¹Pretend that Ralph only has beliefs about himself and Ortcutt.

¹²However, both my account and Lewis's descriptivist account might face a limitation in this vicinity. For neither account can accommodate cases in which Ralph believes that Ortcutt has property F and property G , where F and G are necessarily incompatible in the sense that no individual has both in a single possible world. An example might be something like this: Ralph thinks that Ortcutt is both a human and a robot, not realizing that nothing can be both. But then no possible individual has all the properties Ralph attributes to Ortcutt. I do not know how to solve this problem, and so leave it for future discussion.

that a single actual individual may correspond to two different individuals in one of the agent's doxastic alternatives, for the agent may think 'one is two'. But so far in our development of the sequenced worlds account, we haven't made any allowances for this. Remedyng this is not difficult, but it does require us to revise our account slightly.

Consider the Ortcutt case. In giving a description of that case, it is normal to say something like this:

Ralph is acquainted with Ortcutt in two different ways: he sees Ortcutt giving a rousing speech on the stump, and he later sees Ortcutt sneaking around in a trenchcoat. He doesn't realize that the man giving the speech is the man sneaking around in a trenchcoat. In virtue of being acquainted with Ortcutt in the 'speech' way, he believes that Ortcutt is not a spy. But in virtue of being acquainted with Ortcutt in the 'trenchcoat' way, he believes that Ortcutt is a spy.

The idea here is that Ralph doesn't just believe things about Ortcutt *simpliciter*; he always believes them *relative to* some acquaintance relation or other. If you reflect on the example, this notion is intuitive: when Ralph sees Ortcutt sneaking around in a trenchcoat and thinks *that guy is a spy*, we can say that he believes that Ortcutt is a spy, *relative to* the relation x bears to y just in case x sees y sneaking around in a trenchcoat. When Ralph sees Ortcutt giving a speech on the stump and thinks *he's no spy*, we can say that Ralph believes that Ortcutt is not a spy, *relative to* the relation x bears to y just in case x sees y giving a speech on the stump. The notion of believing something relative to an acquaintance relation is easy to understand, if difficult to analyze.

In specifying the content of Ralph's belief state, I propose to use locutions like 'Ralph believes that Ortcutt is F ', relative to acquaintance relation R '. I will depend on our intuitive understanding of that way of speaking. And I will abbreviate that sort of claim by saying things like 'Ralph believes that Ortcutt- R is F '. But don't take 'Ortcutt- R ' to denote some sort of spooky entity; it merely helps to abbreviate the longer locution.

Let ' R_s ' denote the relation of acquaintance Ralph bears to Ortcutt when Ralph hears Ortcutt deliver the rousing speech, and let ' R_t ' denote the relation of acquaintance Ralph bears to Ortcutt when Ralph sees Ortcutt sneaking around in a trenchcoat. Suppose for simplicity that Ortcutt only has beliefs about himself, about Ortcutt- R_s , and about Ortcutt- R_t .

3. Imagination, Sequenced Worlds, and the *De Re*

That is, suppose there are only three ‘vivid characters’ in Ralph’s ‘inner story’ (cf. Kaplan, 1968), one of whom corresponds to Ralph, two of whom correspond to Ortcutt. In that case, in order to represent Ralph’s doxastic alternatives, we must use sequenced worlds with sequences of length three, rather than sequenced worlds with sequences of length two (which is what we’ve been using up until now). To characterize Ralph’s beliefs we can say that a sequence world $\langle w, \langle x, y, z \rangle \rangle$ is compatible with what Ralph believes about $\langle \text{Ralph}, \text{Ortcutt}, \text{Ortcutt} \rangle$ only if:

- y gave a rousing speech in w ;
- y is not a spy in w ;
- z sneaks around in a trenchcoat in w ; and
- z is a spy in w .

Note that in this specification of what $\langle w, \langle x, y, z \rangle \rangle$ must be like in order to be compatible with what Ralph believes about $\langle \text{Ralph}, \text{Ortcutt}, \text{Ortcutt} \rangle$, we let the second member of the sequence y represent what Ralph believes about Ortcutt relative to R_s , and we let the third member of the sequence z represent what Ralph believes about Ortcutt relative to R_t .

Thus, our specification of what it is for a sequenced world $\langle w, \langle x, y, z \rangle \rangle$ to be compatible with what Ralph believes about $\langle \text{Ralph}, \text{Ortcutt}, \text{Ortcutt} \rangle$ has been implicitly relativized to a sequence of acquaintance relations, in this particular case the sequence is: $\langle \text{identity}, R_s, R_t \rangle$. (I’ll discuss what the identity relation is doing in that sequence in a moment.) The sequence of acquaintance relations tells us which member of the representing sequence represents what Ralph believes about Ortcutt- R_s (in this case, the second) and which represents what Ralph believes about Ortcutt- R_t (in this case, the third). There are two conditions on a sequence of acquaintance relations if it is to do this job: first, it must be of precisely the same length as the *res* sequence $\langle \text{Ralph}, \text{Ortcutt}, \text{Ortcutt} \rangle$ (so in this case, it must be of length three); and second, Ralph must bear the n th acquaintance relation in $\langle \text{identity}, R_s, R_t \rangle$ to the n th member of $\langle \text{Ralph}, \text{Ortcutt}, \text{Ortcutt} \rangle$. From now on we will make explicit the role that a sequence of acquaintance relations is playing in our characterization of doxastic compatibility.¹³

¹³Note that relativity to a sequence of acquaintance relations is only really needed when the same individual appears more than once in the *res* sequence, i.e. in cases where the agent thinks one is two.

Note that a sequenced world compatible with what Ralph believes about $\langle \text{Ralph}, \text{Ortcutt}, \text{Ortcutt} \rangle$ relative to $\langle \text{identity}, R_s, R_t \rangle$ will not be compatible with what Ralph believes about $\langle \text{Ralph}, \text{Ortcutt}, \text{Ortcutt} \rangle$ relative to $\langle \text{identity}, R_t, R_s \rangle$ (note that R_s and R_t are in different places in these two sequences of acquaintance relations). This is because Ralph believes different things about Ortcutt- R_s and Ortcutt- R_t . When saying whether or not a sequenced world is compatible with what Ralph believes about $\langle \text{Ralph}, \text{Ortcutt}, \text{Ortcutt} \rangle$, it is arbitrary which of these two sequences of acquaintance relations we pick: either one will allow to characterize Ralph's belief state. In general, any sequence of acquaintance relations will be suitable so long as it meets the two conditions specified above: (i) the sequence of acquaintance relations is the same length as the *res* sequence $\langle \text{Ralph}, \text{Ortcutt}, \text{Ortcutt} \rangle$, and (ii) Ralph bears the n th acquaintance relation in the sequence of acquaintance relations to the n th member of $\langle \text{Ralph}, \text{Ortcutt}, \text{Ortcutt} \rangle$.

Let me suggest that we adopt a convention concerning the representation of *de se* belief. Our convention is that the agent himself will always be the first member of the *res* sequence, and the relation of identity will always be the first member of the sequence of acquaintance relations. Recall Lewis's idea that *de se* belief is a *de re* belief where the relation of acquaintance is identity (§2.6); we adopt that idea here. So if, in the above example, Ralph believes (*de se*) that he will live a long and prosperous life, then all the sequenced worlds $\langle w, \langle x, y, z \rangle \rangle$ compatible with what Ralph believes about $\langle \text{Ralph}, \text{Ortcutt}, \text{Ortcutt} \rangle$ relative to $\langle \text{identity}, R_s, R_t \rangle$ will be such that x leads a long and prosperous life in w .

So the relation of doxastic compatibility now involves four *relata*: the agent (or agent-at-world); a sequenced world; a *res* sequence; and a sequence of acquaintance relations. For the example above, we can say that a sequenced world $\langle w, \langle x, y, z \rangle \rangle$ is compatible with what Ralph believes about $\langle \text{Ralph}, \text{Ortcutt}, \text{Ortcutt} \rangle$ relative to $\langle \text{identity}, R_s, R_t \rangle$ iff:

- x in w has all the properties Ralph attributes to himself;
- y in w has all the properties Ralph attributes to Ortcutt relative to R_s ;
- z in w has all the properties Ralph attributes to Ortcutt relative to R_t ; and
- x, y , and z in w stand in all the relations that Ralph believes he,

3. Imagination, Sequenced Worlds, and the *De Re*

Ortcutt- R_s , and Ortcutt- R_t stand in.¹⁴

We can understand the notion of content as we did before: simply as a pair of a sequenced world and a set of sequenced worlds. But we now say that the agent holds a belief with a certain content *relative to* a sequence of acquaintance relations. For example, when Ralph believes in w that Ortcutt- R_t is a spy, the content of Ralph's belief is:

$$\langle \langle w, \langle \text{Ralph, Ortcutt, Ortcutt} \rangle \rangle, \{ \langle w', \langle x', y', z' \rangle : z' \text{ is a spy in } w' \} \rangle$$

and he believes this content relative to $\langle \text{identity, } R_s, R_t \rangle$. Ralph's belief is *true* just in case $\langle w, \langle \text{Ralph, Ortcutt, Ortcutt} \rangle \rangle \in \{ \langle w', \langle x', y', z' \rangle : z' \text{ is a spy in } w' \}$, i.e. just in case Ortcutt is a spy in w .

Cases where an agent thinks two are one can be given a similar treatment. Recall the Sparky-Barky case of §2.2: I see Sparky run into my office on Tuesday, and then see Barky run into my office on Friday, and I come to believe that the two dogs are one, i.e. that Sparky is Barky. For simplicity, suppose that I only have beliefs about myself and these two dogs. How do we represent one of my doxastic alternatives? Although there are only two vivid characters in my inner story, we nevertheless use sequenced worlds of length *three* to represent those alternatives. This is so we can represent the fact that there are *in fact* two different dogs that I think are one. Here's how this works: we can say that a $\langle w, \langle x, y, z \rangle \rangle$ is compatible with what I believe about $\langle \text{Dilip, Sparky, Barky} \rangle$ (relative to the appropriate sequence of acquaintance relations) only if:

- $y = z$ in w ; and
- y is a friendly dog in w .

So although we use sequenced worlds of length three, the fact that I think Sparky is Barky is represented in the theory by the fact that the individual that represents Sparky is identical to the individual that represents Barky.

De se identity puzzles can be given a parallel treatment. Recall Situation 4 from §1.3, where amnesiac Lingens is reading *The Life of Rudolph Lingens* and becomes convinced that he is not Lingens. Suppose that Lingens only has *de se* beliefs about himself and ‘third-personal’ *de re* beliefs

¹⁴Here you can see clearly that in order to understand this notion of doxastic compatibility, we need to rely on our intuitive understanding of the idea that an agent has beliefs about a *res* relative to an acquaintance relation.

about himself in virtue of bearing the acquaintance relation R_r to himself (R_r = the relation x bears to y just in case x reads a biography of y), so that Lingens's inner story only contains two vivid characters. We can represent one of Lingens's doxastic alternatives using a sequenced world of length two. We can say that a sequenced world $\langle w, \langle x, y \rangle \rangle$ is compatible with what Lingens believes about $\langle \text{Lingens}, \text{Lingens} \rangle$ relative to $\langle \text{identity}, R_r \rangle$ only if:

- x is a gentle soul in w ; and
- y is a vicious character in w .

Assuming no gentle soul is also a vicious character, it follows from the above that $x \neq y$ in w , which reflects the fact that Lingens believes he is not Lingens- R_r .

3.3 Generalizing the theory

So far, we've been discussing toy cases in which an agent only has beliefs about himself and one or two other individuals. But of course in any realistic case, an agent will have beliefs about many more individuals. The extension of our theory to such cases is straightforward.

A sequenced world can now be taken to be a pair consisting of a world and a sequence of n individuals. Let ' x_N ' denote the sequence $\langle x_1, \dots, x_n \rangle$, and similarly for ' y_N ', ' x'_N ', and so on. Suppose an agent x has beliefs about n individuals in w . The pair $\langle y_N, R_N \rangle$ is an *acquaintance pair* for x in w iff:

- y_N is a *res* sequence for x in w , i.e. it is a sequence consisting of all and only those individuals that x is acquainted with in w , and $y_1 = x$;
- for all $m \leq n$, x bears R_m to y_m in w , and R_1 is the identity relation.

If $\langle y_N, R_N \rangle$ is an acquaintance pair for x in w , then a sequenced world $\langle w', y'_N \rangle$ is compatible with what an agent x believes in w about y_N relative to R_N iff the y'_i stand in every n -ary relation \mathcal{R} that x believes the $y_i \cdot R_i$ stand in. (Recall our earlier explanation of abbreviations like 'Ortcutt- R '.) Note that since any m -ary relation ($m \leq n$) can be represented as an n -ary relation, this gloss on the notion of compatibility should suffice.

3. Imagination, Sequenced Worlds, and the *De Re*

The *content* of a belief is a pair $\langle\langle w, y_N \rangle, p \rangle$ of a sequenced world and a set of sequenced worlds, where w is the world of the agent who holds the belief (that agent being y_1), y_N is a *res* sequence for y_1 in w , and p is a set of sequenced worlds. Contents are believed relative to a sequence of acquaintance relations. If $\langle y_N, R_N \rangle$ is an acquaintance pair for y_1 in w , then y_1 believes $\langle\langle w, y_N \rangle, p \rangle$ relative to R_N iff all the sequenced worlds $\langle w', y'_N \rangle$ compatible with what y_1 believes in w relative to R_N are contained in p , i.e. y_1 locates $\langle w, y_N \rangle$ in p . If an agent has a belief with content $\langle\langle w, y_N \rangle, p \rangle$, that belief will be *true* just in case $\langle w, y_N \rangle \in p$.

We can give a more precise account of *de re* belief now. The account goes like this:

Sequenced *De Re* Belief

x believes (*de re*) in w that z is F iff for any acquaintance pair $\langle y_N, R_N \rangle$ for x in w , there is an $m \leq n$ such that $y_m = z$, and every sequenced world $\langle w', y'_N \rangle$ compatible with what x believes in w about y_N relative to R_N is such that y'_m is F in w'

Note that if $\langle y_N, R_N \rangle$ is an acquaintance pair for x in w , and there is an $m \leq n$ such that $y_m = z$, it follows (by the definition of an acquaintance pair) that x bears R_m to z in w . Thus, it follows that x has a *de re* belief about z in w only if x is acquainted with z in w . Note also that, on this account, *de se* belief comes out as a special case of *de re* belief: where the *res* is the agent, and the acquaintance relation is identity.

3.4 Imagining and sequenced worlds

So far we have only discussed the sequenced worlds account of *de se* and *de re* belief. But what about counterfactual attitudes like imagining? How are they to be treated in this framework? The treatment of ‘third-personal’ *de re* imagining is straightforward, and requires no amendments to the above account: we simply take our sequenced worlds account of belief and replace the word *believes* with the word *imagines*. The resulting theory has no trouble with third-personal *de re* imaginings.

Consider one of the cases that was a problem for General Lewis (§2.3): Ralph imagines that he is not acquainted with Ortcutt. Let R be the acquaintance relation Ralph bears to Ortcutt, and let’s suppose that that Ralph is imagining a situation in which only he and Ortcutt exist, but are not acquainted with one another. On our account, this means that all the sequenced worlds $\langle w, \langle x, y \rangle \rangle$ compatible with what Ralph imagines about

$\langle \text{Ralph, Ortcutt} \rangle$ relative to $\langle \text{identity, } R \rangle$ are such that x is not acquainted with y in w . And there are certainly sequenced worlds that meet that condition, so the content of the imagining will not be empty.

The problem of counterfactual attitudes arose for General Lewis because on that theory, one identifies Ortcutt's representative in one of Ralph's imagination alternatives via the acquaintance relation that the center of the alternative bears to that representative. And that means that the theory cannot handle cases in which Ralph imagines not being acquainted with Ortcutt. But the sequenced worlds account has a different method of picking out the individual who represents Ortcutt in one of Ralph's imagination alternatives: the method of stipulation. We don't require that a sequenced world $\langle w, \langle x, y \rangle \rangle$ must be such that x is acquainted with y in w in order for it to be compatible with what Ralph imagines about $\langle \text{Ralph, Ortcutt} \rangle$. So there is no trouble characterizing the content of *de re* imaginings on the sequenced worlds account.

What about *de se* imaginings? How does the sequenced worlds account help with the problem of the impersonal-*de se*? There is a *prima facie* difficulty here. As the theory has been developed thus far, we only have one 'center': the first member of the sequence in a sequenced world is always stipulated to represent the agent relative to the relation of identity. But in order to solve the problem of the impersonal-*de se*, we need *two* centers. To get around this, we can add another 'center' to the imagination-alternatives, and stipulate that one center is the 'inside center' and one the 'impersonal-*de se* center'. The idea would be to understand imagination alternatives as triples $\langle w, x, y_N \rangle$, with x the inside center and y_1 the impersonal-*de se* center.¹⁵

On the envisioned account, when I imagine that I am Brigitte Bardot and that I am kissing me, the content of my imagining is:

$$\langle \langle \text{actual world}, \langle \text{Dilip, Bardot} \rangle \rangle, \{ \langle w, x, \langle y_1, y_2 \rangle \rangle : x \text{ is } y_2 \text{ in } w \text{ and } x \text{ is kissing } y_1 \text{ in } w \} \rangle$$

On this account, impersonal-*de se* imagining will count as a species of *de re* imagining about oneself where the relation of acquaintance is identity.

¹⁵Instead of the 'extra center' approach, we could have gone two-dimensional here as well. This time we would have used functions from sequenced belief worlds to centered propositions, rather than two-dimensional centered intensions, to help characterize the content of an imagining. But this strikes me as a less attractive proposal than the one I'm about to offer, in part because here the two-dimensional move is not needed in order to characterize the contents of third-personal *de re* imaginings.

3. Imagination, Sequenced Worlds, and the *De Re*

But on this account, inside imagining is not a kind of *de re* imagining at all, since the inside center *x* does not represent anyone in the *res* sequence.

This feature of the theory might seem odd: after all, when I imagine something from the inside, am I not imagining something about *me*? Shouldn't such imaginings thus count as *de re* imaginings about myself? That might seem right, but not everyone agrees with this claim: several philosophers have claimed that when I imagine something from the inside I am *not* (or not necessarily) imagining anything about the person I in fact am. For example, when I imagine from the inside being Napoleon, my imagining is not in any sense ‘about’ Dilip. Here is David Velleman:

If my approach to imagining that I am Napoleon... is to imagine *being* Napoleon, then I simply imagine a particular experience as experienced *by* Napoleon. I imagine the landscape as seen through Napoleon’s eyes, the sounds of battle as heard through his ears, the nap of a tunic as felt by his hand. Although Napoleon doesn’t appear in the resulting mental image, he does appear in the content of my imagining, since I am imagining Austerlitz as experienced by him. But I, David Velleman, am absent both from the image and from the content of the imagining: I’m not imagining anything about the person who I actually am. (Velleman, 1996, 40)

Williams (1966, 40-45) and Reynolds (1989) reach similar conclusions. This (admittedly somewhat elusive) feature of imagining from the inside is reflected to some extent in our theory, since, according to it, the inside center does not represent anyone in the agent’s *res* sequence.

Using the appropriate notions of ‘interesting’ and ‘boring’ sequenced propositions (sets of sequenced worlds), we can characterize outside, inside, and impersonal-*de se* imaginings. I trust the reader can see how to extend the boring-interesting technique of §1.2.1 and §2.5 to this theory as well, and so I will not go through that exercise here.

3.5 Attitude ascription again

The semantics that accompanies the sequenced worlds account differs only minimally from the semantics of §2.6. We retain the same basic semantic framework of §1.2, as well as the account of *de se* ascription developed there.

Let's recall what we want the truth conditions of a *de re* belief ascription to be. Consider:

- (9) Ralph believes that Ortcutt is a spy.

This should be true at a world w just in case any acquaintance pair $\langle y_N, R_N \rangle$ for Ralph in w is such that there is an $m \leq n$ such that $y_m =$ Ortcutt, and every sequenced world $\langle w', y'_N \rangle$ compatible with what Ralph believes in w about y_N relative to R_N is such that y'_m is a spy in w' .

To derive this result, let us keep the LF of (9) that we had earlier in §2.6:

- (10) Ralph [[believes Ortcutt] [$\lambda_1 \lambda_2 t_2$ is a spy]]

And let us adopt the following lexical entry for *believes*:

$\llbracket \text{believes} \rrbracket^{c,i,g} = \lambda res_e. \lambda p_{\langle s, \langle e, et \rangle \rangle}. \lambda att_e.$ for any acquaintance pair $\langle y_N, R_N \rangle$ for att in w_i , there is an $m \leq n$ such that $y_m = res$, and every sequenced world $\langle w', y'_N \rangle$ compatible with what att believes in w_i about y_N relative to R_N is such that $p(w')(y'_1)(y'_m) = 1$.¹⁶

With those assumptions in place, we compute the truth conditions of (9) as follows:

$\llbracket \text{Ralph believes Ortcutt } \lambda_1 \lambda_2 t_2 \text{ is a spy} \rrbracket^{c,i,g} = 1$ iff

For any acquaintance pair $\langle y_N, R_N \rangle$ for Ralph in w_i , there is an $m \leq n$ such that $y_m =$ Ortcutt, and every sequenced world $\langle w', y'_N \rangle$ compatible with what Ralph believes in w_i about y_N relative to R_N is such that $\lambda w. \llbracket \lambda_1 \lambda_2 t_2 \text{ is a spy} \rrbracket^{c,w,g}(w')(y'_1)(y'_m) = 1$ iff

For any acquaintance pair $\langle y_N, R_N \rangle$ for Ralph in w_i , there is an $m \leq n$ such that $y_m =$ Ortcutt, and every sequenced world $\langle w', y'_N \rangle$ compatible with what Ralph believes in w_i about y_N relative to R_N is such that y'_m is a spy in w' .

¹⁶Note that we're universally quantifying over acquaintance pairs. Alternatively, we could have had the model specify a unique acquaintance pair for each agent x at world w . Nothing seems to hinge on this choice.

Let me also sketch a lexical entry for *imagines*. The difference between the entry for *imagines* and the one for *believes* is minor: the only difference is in the sort of sequenced world quantified over. In the case of *imagines*, sequenced worlds have an extra center, and so sequenced worlds are triples rather than pairs. Here is the entry:

$\llbracket \text{[imagines]} \rrbracket^{c,i,g} = \lambda res_e. \lambda p_{\langle s, \langle e, et \rangle \rangle}. \lambda att_e.$ for any acquaintance pair $\langle y_N, R_N \rangle$ for att in w_i , there is an $m \leq n$ such that $y_m = res$, and every sequenced world $\langle w', x', y'_N \rangle$ compatible with what att imagines in w_i about y_N relative to R_N is such that $p(w')(x')(y'_m) = 1$.

It is easy to confirm that this account has no problem with counterfactual attitude ascriptions like (11):

- (11) Ralph imagined that he had never been acquainted with Ortcutt.

As before, the structure of this sentence is:

Ralph [[imagined Ortcutt] $[\lambda_1 \lambda_2 \text{he}_1 \text{ had never been acquainted with } t_2]$]

And with this structure and lexical entry, we can compute the truth conditions of (11) as follows:

$\llbracket \text{[Ralph imagined Ortcutt } \lambda_1 \lambda_2 \text{ he}_1 \text{ had never been acquainted with } t_2] \rrbracket^{c,i,g} = 1 \text{ iff}$

For any acquaintance pair $\langle y_N, R_N \rangle$ for Ralph in w_i , there is an $m \leq n$ such that $y_m = \text{Ortcutt}$, and every sequenced world $\langle w', x', y_N \rangle$ compatible with what Ralph imagines in w_i about y_N relative to R_N is such that $\lambda w. \llbracket \lambda_1 \lambda_2 \text{ he}_1 \text{ had never been acquainted with } t_2 \rrbracket^{c,w,g}(w')(x')(y'_m) = 1$ iff

For any acquaintance pair $\langle y_N, R_N \rangle$ for Ralph in w_i , there is an $m \leq n$ such that $y_m = \text{Ortcutt}$, and every sequenced world $\langle w', x', y_N \rangle$ compatible with what Ralph imagines in w_i about y_N relative to R_N is such that x' is not acquainted with y'_m in w' .

As with the semantics in §2.6, impersonal-*de se* imaginings are simply a species of *de re* imagining, and so are reported by *de re* imagination reports. To see this, recall Situation 1 from §1.3: amnesiac Lingens imagines from the outside that he is skiing down a steep hill. This is an impersonal-*de se* imagining. To report this imagining, one might use (12):

- (12) Lingens imagined that he was skiing down a steep hill.

On our semantics, this sentence has two structures: one in which *he* is bound by the attitude verb, one in which it is not (cf. §2.6). The former gives rise to the inside reading, and so is not true in Situation 1. The latter gives rise to the outside *de re* reading. Assuming the variable assignment assigns Lingens to *he*, then (12) is true at a world *w* iff:

For any acquaintance pair $\langle y_N, R_N \rangle$ for Lingens in *w*, there is an $m \leq n$ such that $y_m = \text{Lingens}$, and every sequenced world $\langle w', x', y'_N \rangle$ compatible with what Lingens imagines about y_N relative to R_N is such that y'_m is skiing down a steep hill in *w'*.

Note that Lingens impersonally-*de se* imagines in a world *w* that he is skiing down a steep hill iff

For any acquaintance pair $\langle y_N, R_N \rangle$ for Lingens in *w*, $y_1 = \text{Lingens}$, and every sequenced world $\langle w', x', y'_N \rangle$ compatible with what Lingens imagines about y_N relative to R_N is such that y'_1 is skiing down a steep hill in *w'*.

This means that, if Lingens impersonally-*de se* imagines skiing down a steep hill in *w*, then for any acquaintance pair $\langle y_N, R_N \rangle$ for Lingens in *w*, there will be an $m \leq n$ that meets the appropriate conditions for (12)'s truth condition to be satisfied: $m = 1$. So if Lingens impersonally-*de se* imagines that he is skiing down a steep hill, he imagines *de re* that he is skiing down a steep hill, and so (the free structure of) (12) is true.

Chapter 4

Imagination, Possibility, and Personal Identity

The previous three chapters have all dealt with issues concerning how to represent the content of various kinds of imagining. In this chapter, we turn our attention to the role of imagination in metaphysical thought experiments, and, in particular, personal identity thought experiments. But issues of content will re-surface: one of the claims of this chapter is that being clear about the contents of inside vs. outside imaginings leads us to a conception of possibility which helps account for some otherwise puzzling intuitions about personal identity over time.

4.1 Introduction

A familiar divide in the theory of personal identity over time is between those who think personal identity is essentially a matter of psychological continuity and those who think it is essentially a matter of bodily continuity. But proponents of these two views share a common opponent: the philosopher who thinks that personal identity is a ‘further fact’, something over and above facts about impersonal continuity relations. The view that facts about personal identity are metaphysically independent of facts about continuity is often associated with substance dualism, but it is not immediately clear whether or not the view requires dualism. Following Parfit (1984), let us call this view, whether dualist or not, *the Simple View*, and let us call the view that personal identity is essentially a matter of some kind of impersonal continuity *the Complex View*. Since most philosophers who have thought about the matter take it that personal identity consists in psychological or bodily or perhaps brain continuity, most philosophers who have thought about the matter endorse the Complex View.

A natural way to try to characterize the Simple-Complex dispute is

to understand it as a dispute over whether or not personal identity *supervenes on* lower-level continuity relations, like psychological, bodily, and brain continuity. Does fixing all the facts about psychological, bodily, and brain continuity fix the facts about personal identity? The Complex View maintains, and the Simple View denies, that personal identity facts are fixed by continuity facts.

Advocates of the Simple View defend their position by noting that when one looks at the standard personal identity puzzle cases from the first-person point of view (or from the inside), it seems that personal identity does not supervene.¹ I think this is right, and I think advocates of the Complex View have failed to appreciate the significance of this point. But, as we will see, this is only half the story; for when we look at the same hypothetical cases from the third-person perspective (or from the outside), it seems that personal identity *does* supervene. I think advocates of the Simple View have failed to appreciate the significance of *this* point. Note the puzzle that emerges: from the third-person point of view, personal identity seems to supervene, but from the first-person point of view, it seems not to supervene. My aim in this chapter is to examine, and then solve, this puzzle.

The chapter is structured as follows. In the next section, I try to bring the Simple-Complex dispute into sharper focus; this involves a brief discussion of some background issues in the metaphysics of persistence. In §4.3, I discuss the puzzle as it arises in the context of two well-known thought experiments: fission and ‘body-switching’. In §4.4, I develop a picture of imagination and possibility that I think is needed to solve the puzzle. The imagination part of the picture will be familiar from Chapter 1. What’s added here is the suggestion that we take seriously the idea that corresponding to these two forms of imagination are two kinds of metaphysical possibility, one analyzed (as is familiar) in terms of quantification over possible worlds, the other – and this is less familiar – in terms of quantification over centered worlds (cf. Lewis, 1983a, 1986, §4.4). The upshot is that personal identity *does* supervene on continuity, even though there is a perfectly good and important sense in which settling all the continuity facts does not settle my future (§§4.5 - 4.6).

¹For example, see Chisholm (1969), Swinburne (1984), and Baker (2000) (though these authors don’t put the point in terms of supervenience).

4.2 Supervenience

I said that it was natural to try to characterize the Simple-Complex dispute as a dispute over whether or not personal identity supervenes on lower-level continuity relations, like psychological, bodily, and brain continuity. But just how should we state the relevant supervenience thesis? In stating the supervenience thesis, I'm going to employ the *four-dimensionalist's* language of *person stages* and *temporal parts*. Four-dimensionalism (or *perdurantism*) is the view that persisting things like persons, rocks, and ships are four-dimensional objects spread out in time and space, things with temporal as well as spatial parts (see e.g. Lewis, 1976). Not everyone is a four-dimensionalist: *three-dimensionalists* (alternatively: *endurantists*) reject the idea that ordinary objects have temporal parts (see e.g. van Inwagen, 1990). On their view, an object that exists at t is wholly present at t , with no parts existing at other times.²

Given the controversy between these two camps, it might be nice to state our supervenience thesis without making it look as though we were taking sides on this issue. But it will be much easier if we simply help ourselves to the language of four-dimensionalism for the moment, and then later return to say something about how these matters would look through a three-dimensionalist lens. This is easier because four-dimensionalism allows for a simple and intuitive characterization of the issues at hand. So I will use the language of four-dimensionalism throughout most of the chapter, and then return in §4.5.4 to the question of how things look ‘in 3D’.³

Back to supervenience. Let us say that two pairs of possible person stages $\langle x, y \rangle$ in w and $\langle x', y' \rangle$ in w' are *the same with respect to continuity* iff:

- x is psychologically continuous with y in w iff x' is psychologically continuous with y' in w' ;
- x is bodily continuous with y in w iff x' is bodily continuous with y' in w' ;

²For a thorough discussion of how to characterize these two views, see Sider (2001b, Ch.3).

³A note on terminology: sometimes I use *person stage* and *temporal part* to refer to more-or-less instantaneous ‘slices’ of persons; other times, I use them to refer to aggregates of instantaneous stages, i.e. ‘thick slices’ of persons. (On the latter usage, persons are improper temporal parts of themselves.) Context should disambiguate.

- x is ‘brain continuous’ with y in w iff x' is ‘brain continuous’ with y' in w' ;
- and so one for each relevant continuity relation.

Then we can state our supervenience thesis as follows.

Supervenience

For all worlds w, w' , and pairs of person stages $\langle x, y \rangle$ in w , $\langle x', y' \rangle$ in w' : if x and y in w are the same with respect to continuity as x' and y' in w' are, then x and y are stages of the same person in w just in case x' and y' are stages of the same person in w' .

Note that the psychological, bodily, and brain continuity theories of personal identity all entail **Supervenience**, which suggests that **Supervenience** captures what is common to the various versions of the Complex View. For example, proponents of the Psychological Continuity Theory argue for the following criterion of personal identity:

Necessarily, for all x, y : x and y are stages of the same person just in case x and y are related by the relation of psychological continuity.⁴

This theory identifies the personal identity relation with the relation of psychological continuity. The Bodily Continuity Theory, on the other hand, identifies the personal identity relation with the relation of bodily continuity. Since both these relations are in the relevant supervenience base, both theories entail **Supervenience**. This suggests that it makes sense to characterize the Complex View as the acceptance of **Supervenience**, and to characterize the Simple View as the denial of **Supervenience**.⁵

⁴What is the relation of psychological continuity? Two person stages x and y are *psychologically connected* iff x and y have psychological states (beliefs, desires, intentions, apparent memories, character traits, etc.) which are (a) similar in content, and (b) causally connected in the right way, i.e. the psychological states of the later person stage causally depend (in the right way) for their character on the states of the earlier one (Lewis, 1976, 55-56). *Psychological continuity* is then the ancestral of psychological connectedness.

⁵One problem with characterizing the Simple View as the denial of **Supervenience** is that, as Noonan (1989, 96 - 97) notes, most advocates of the Simple View defend a stronger thesis: they defend the claim that there are no non-trivial necessary conditions on personal identity. For example, they would deny the claim that, necessarily, x and

4.3 The puzzle

4.3.1 Fission

Suppose that we could divide my brain, and transplant each hemisphere into a different body (suppose these bodies are qualitatively identical). And suppose that the two resulting persons both had beliefs, desires, intentions, apparent memories, and character traits which were similar to mine. In short, suppose that both resulting persons were psychologically continuous with me. This is a fission case, impersonally described.

In the impersonal fission case, most people are inclined to say that the initial person does not survive fission. The reason for this is that, in order for the initial person to end up as, say, the left hemisphere person (*Lefty*), there would need to be some fact that explained her ending up as Lefty rather than as the right hemisphere person (*Righty*).⁶ But since the initial person bears exactly the same (qualitative) relations to each of Lefty and Righty, there can be no such fact. This suggests that the initial person cannot be Lefty, nor can she be Righty. And since Lefty and Righty are distinct, the initial person cannot be identical to *both* Lefty and Righty. The only option left is that the initial person does not survive.

If this is right, then advocates of the Psychological Continuity Theory would need to revise their account so it includes a ‘non-branching’ clause. The relation of non-branching psychological continuity obtains between two person stages x and y just in case (i) they stand in the relation of psychological continuity, and (ii) there is no person stage z ($z \neq x$ and $z \neq y$) such that one and only one of x and y stands in the relation of psychological continuity to z .⁷ Similar revisions could be made to the Brain and Bodily Continuity Theories. In addition to psychological, bodily,

y are stages of the same person only if x and y have the same body. Thus, there is a possible intermediate position between the Simple View and the Complex View: one which denies *Supervenience* but maintains that there are some non-trivial necessary conditions on personal identity. As Noonan notes, this view is not popular, and is not particularly plausible. (This contrasts with the case of trans-world identity where the analogous position is the standard one.) Like Noonan, I will ignore this intermediate position: thus, if a consideration speaks in favor of the existence of a non-trivial necessary condition on personal identity, I take it to speak in favor of *Supervenience*.

⁶Note that *Lefty* and *Righty* are not names but abbreviations of the definite descriptions *the person with the left hemisphere* and *the person with the right hemisphere* respectively.

⁷There may be a worry about circularity in this statement of the non-branching criterion; see Thomson (1997, 214-15) and Brueckner (2005) for discussion.

and brain continuity, we can now think of Supervenience as including non-branching versions of those relations in the supervenience base.

If this is our reaction to the impersonally described fission case, then our reaction seems to support Supervenience.⁸ For Supervenience leads us to expect that, once the continuity relations are fixed as they are in the fission case, there will not be multiple ‘person possibilities’ compatible with our description of the case. And here we find this expectation met, for once we fix the continuity facts in the fission case, we seem to see only one person possibility: the initial person does not survive. If the continuity facts did not fix the person facts, we might expect there to be *three* possibilities for the initial person’s survival: she ends up as Lefty, she ends up as Righty, or she doesn’t end up at all. That we think the latter is the only possibility consistent with the specified continuity facts suggests that we think that once the continuity facts are fixed as they are in the fission case, the relevant person facts are also fixed.

There is a great deal of literature on the fission case, and not everyone is happy with the above solution, though I think that solution has some claim to being the dominant response in the literature. Two other responses are worth mentioning:

- It is indeterminate whether the initial person survives. (Parfit, 1984)
- There are two persons in a fission case, and they share their birth-to-fission temporal part. (Lewis, 1976)

The important issue for us is not to decide on which of these three views is right. The important point for us is to note that all three of these responses seem to support the idea that fixing the continuity facts fixes the person facts.⁹ For none of these views holds that there are multiple ways things might turn out for the initial person(s). All agree that there is only one person possibility consistent with the specified continuity facts; they simply disagree about what that person possibility is. So although it may not be clear how best to describe the fission case when we consider it from the third-person perspective, it seems that looking at the case from

⁸Remember, as I noted earlier (see footnote 5), if a case suggests that there is a non-trivial necessary condition on personal identity, I take it to support Supervenience.

⁹Making this point precisely with respect to the indeterminacy-view might require us to re-formulate Supervenience along the following lines: If two pairs of possible person stages are the same continuity-wise, then either (i) both pairs are determinately person-related; (ii) both pairs are determinately not person-related; or (iii) both pairs are such that it is indeterminate whether or not they are person-related.

4. Imagination, Possibility, and Personal Identity

this perspective supports *Supervenience*. For ease of exposition I will for the most part simply assume that the first response to the fission case – according to which the initial person does not survive – is the intuitively right one, but nothing of substance hangs on this choice. All that really matters is the claim that our reaction to the impersonally-described fission case supports *Supervenience*.

The above case was specified impersonally: even though I stipulated that it was *my* brain that was divided, we were not called on to take up my point of view or that of any of the participants in the case. We could just have well described the case by saying, *Suppose a person's brain was divided....* But now consider what happens when we do describe the case from the point of view of one of the participants, and imagine following that point of view through time. As several authors have noted, when we do this, we can imagine surviving fission as either one of the resulting persons.¹⁰ Looked at from the first-person point of view, I can see three possibilities for my survival: I survive as Lefty, I survive as Righty, I don't survive at all.

To see this, let us begin with a case of brain division and transplant that does *not* involve fission:

Friday: You have suffered great trauma to your person. The only part of your body that the surgeons believe can be salvaged is the left hemisphere of your brain. The surgeons decide to attempt to transplant the left hemisphere of your brain into another body. The procedure is risky, and you are deeply afraid...

Saturday: Luckily, however, you wake up the next morning, groggy but thrilled to be alive. The surgeons tell you that the transplant was a success. You examine your new body, comparing it favorably with the old one. The surgeons hand you an ice-filled cooler that contains your right hemisphere. You think, *How strange... this is half my old brain.*

Given that we can imagine being in this predicament, it seems that we can imagine surviving fission simply by ‘adding on’ the existence of a second person. Instead of simply waking up, this time I wake up and look over at another person, who I am told has half my old brain:

¹⁰For example: Chisholm (1969, 106), Swinburne (1984, 17-19), Blackburn (1997, 181), and Baker (2000, 137).

Friday: [As above.]

Saturday: Luckily, however, you wake up the next morning, groggy but thrilled to be alive. You examine your new body, comparing it favorably with the old one. The surgeons are pleased—in fact, they are amazed, for not only did you survive, but they somehow managed to successfully transplant your right hemisphere into another body! You catch a glimpse of another person on the right side of the recovery room. You think, *How strange... He has half of my old brain.*

And it seems that one can just as easily imagine winding up as the person on the *other* side of the the recovery room, looking across at the person on the *left* side of the room. A third possibility, of course, is that one can imagine facing the procedure on Friday and then not waking up as either of the resulting persons. Or at least one can imagine this insofar as one can imagine dying.

The first-personally presented fission case seems to tell against Supervenience. For there seem to be three relevant cases, all of which agree on the continuity relations that obtain between me (the initial person), Lefty, and Righty, but which differ on which of those stages are stages of the same person (in one, I end up as Lefty, in other as Righty, in a third, I end up as no one at all).¹¹ Thus, there seem to be *three* possibilities for my persistence which are consistent with the same set of continuity facts. This seems to tell against Supervenience.

This is our puzzle about fission: when we imagine fission from the outside, our judgments support Supervenience; when we imagine it from the inside, our judgments tell against Supervenience.

4.3.2 Williams's puzzle

We now show how this same phenomenon arises in connection with a different hypothetical case: a case of ‘body-switching’.

Williams (1970) invites us to suppose that there is a device capable of ‘extracting’ all or most of the information (beliefs, desires, intentions, apparent memories, character traits, etc.) from a person’s brain. And we are to suppose that the information can then be ‘re-inserted’ back into the

¹¹The initial person, Lefty, and Righty are all aggregates of person stages. When I speak of two aggregates of person stages s and s' being related by a continuity relation R , understand this as saying that any arbitrary instantaneous slice of s bears R to any arbitrary instantaneous slice of s' .

4. Imagination, Possibility, and Personal Identity

person's brain. Imagine two persons, *A* and *B*, each entering a similar machine which 'sucks' all the psychological information out of their brains and puts it into the brain which originally belonged (and may still belong) to the other person, so that, after this procedure, the person in the *A*-body (i.e. *A*'s original body) now has all the memories, thoughts, feelings, etc. that *B* had before the procedure. And similarly, *A*'s psychological information is now in the *B*-body (i.e. *B*'s original body).

The question now is: to whom does each body belong? Does the *B*-body still belong to *B* or does it now belong to *A*? Williams observes that most of us are inclined to say that the *B*-body now belongs to *A*, and that the *A*-body now belongs to *B*. For the person in the *B*-body will have all of *A*'s apparent memories, and none of *B*'s. Further, having *A*'s beliefs, desires, and character traits, the person in the *B*-body will tend to act and talk just as *A* acted and talked. And, of course, this person will think that he is *A*, since he has all of *A*'s beliefs and apparent memories. All this suggests that the person in the *B*-body *is A*. Similar considerations suggest that *B* is the person in the *A*-body after the experiment.

For these reasons, the case, as Williams notes, seems to be one in which two people 'change bodies'. And this seems to show that who a person will be in the future depends on psychological, rather than on bodily or brain, continuity. The point I wish to focus on again is that our initial reaction to this case seems to support **Supervenience**, since **Supervenience** leads us to expect that if we fix the continuity facts in a given situation, then only one 'person possibility' will be compatible with our description of that situation. If the continuity facts did not fix the person facts, there might be two possible situations consistent with Williams's description: one in which the participants switch bodies, another in which they remain in their respective bodies. That we think the latter is not a possibility consistent with the specified continuity facts suggests that we think that once the continuity facts are fixed in the way Williams fixes them, the relevant person facts are also fixed.

So far, the case is not a puzzle—it's simply a case which seems to support the Psychological Continuity Theory over the Simple View and other versions of the Complex View. The puzzle is generated by placing the above description of the case next to a description of the case from the perspective of one of its participants. My description is adapted from Williams (1970, 51-52):

Someone in whose power I am tells me that I am going to be tortured tomorrow. I am frightened, and look forward to to-

morrow in great apprehension. He adds that when the time comes, I shall not remember being told that this was going to happen to me, since shortly before the torture something else will be done to me which will make me forget the announcement. This certainly will not cheer me up. He then adds that when the moment of torture comes, I shall not remember any of the things I am now in a position to remember. This does not cheer me up either. He now further adds that at the moment of torture I shall not only not remember the things that I am now in a position to remember, but will have a different set of impressions of my past, quite different from the memories I have now. I do not think that this would cheer me up either. Nor do I see why I should be put into any better frame of mind by the person in charge adding that the impressions of my past with which I shall be equipped on the eve of torture will exactly fit the past of another person now living. And things would be no better if, finally, he adds that something will happen to that other person so that he will wake up tomorrow unable to remember the thing things he now remembers, and will instead be equipped with impressions of *my* past; and that, far from being tortured, the other person will receive a substantial reward. Fear, surely, would still be the proper reaction: and not because one did not know what was going to happen, but because in one vital respect at least one did know what was going to happen—torture, which one can indeed expect to happen to oneself, and to be preceded by certain mental derangements as well.¹²

Williams then writes:

If this is right, the whole question seems now to be totally mysterious. For what we have just been through is of course merely one side, differently represented, of the transaction which we considered before... (Williams, 1970, 52-53)

Williams points out that the main difference between the two presentations is that the first presentation is conducted entirely in third-personal terms, whereas, in the second presentation, the case is presented from the

¹²This follows Williams's text closely, but is not a quotation. I have amended the case to avoid certain irrelevant complications.

4. Imagination, Possibility, and Personal Identity

first-person perspective—it is presented as happening to *me*. But how does Williams's puzzle constitute an instance of our puzzle? So far we've only seen that imagining this case from the first-person perspective tells against the Psychological Continuity Theory. Since it seems that I would remain in my body after the procedure in this case, that might be because the Simple View is true, but it might be because the Bodily or Brain Continuity Theory is true. The latter is what Williams seems to conclude about this second presentation:

It is often recognized that there are ‘first-personal’ and ‘third-personal’ aspects of questions about persons, and that there are difficulties about the relations between them. It is also recognized that ‘mentalistic’ considerations... and considerations of bodily continuity are involved in questions of personal identity... It is tempting to think that the two distinctions run parallel: roughly, that a first-person approach concentrates on mentalistic considerations, while a third-person approach emphasizes considerations of bodily continuity. The present discussion is an illustration of exactly the opposite. (Williams, 1970, 62)

But a *third* presentation of Williams's case – one which Williams does not consider – shows that this conclusion is too quick. Taken together, the second and third presentations of the case seem to suggest that the ‘first-person approach’ supports the Simple View, not the Bodily or Brain Continuity Theory (cf. Madell, 1981, 94).

The third presentation of the case is simply one in which the experimenter tells me that I will undergo a certain medical procedure which will result in my waking up tomorrow in a different body, with all my memories and other psychological states intact.¹³ The body I will wake up in tomorrow is the body of another person, and that person will wake up with my body tomorrow. The experimenter also tells me that tomorrow I will be tortured, and that the other person will receive a reward. While not considered by Williams, this third presentation is a type of thought experiment very familiar from the personal identity literature: imagine you wake up in a different body tomorrow, with all your memories and psychological states intact. Many have used this sort of case to argue for the Psychological Continuity Theory.¹⁴

¹³I assume, of course, that the medical procedures in all three presentations are exactly the same.

¹⁴For discussion of this sort of case, see Shoemaker (1963) and Thomson (1997, 217-18).

Taken together, the second and third presentations seem to conflict with Supervenience. An implicit assumption of the second and third presentations is that the continuity relations between the relevant person stages are the same in both cases. Let's call the body I start out in in both cases *the A-body*, and the other relevant body in both cases *the B-body*. Let '*preA*' be an abbreviation for *the person in the A-body before the procedure*, and let '*postA*' be an abbreviation for *the person in the A-body after the procedure*, and define '*preB*' and '*postB*' in an analogous fashion. (Perhaps these bodies have large scarlet letters painted on them.) So in both cases, *preA* (me) and *postA* are related by bodily and brain, but not psychological, continuity. But in the second presentation, I end up as *postA* (I remain in my body) whereas in the third I do not, since I end up as *postB* (I go where my psychology goes). So we have two cases in which *preA* and *postA* are the same with respect to continuity, even though the two cases differ over whether *preA* and *postA* are stages of the same person or not. The apparent existence of these two possibilities conflicts with Supervenience.

This is our puzzle about Williams's case: when we imagine the case from the outside, our judgments support Supervenience, and in particular, the Psychological Continuity Theory; but when we imagine it from the inside, our judgments tell against Supervenience.

Note that the majority of commentators on Williams's puzzle seem to have missed what it actually shows. Discussions of this case have tended to follow Williams in taking it to show that we have intuitions that support the Psychological Continuity Theory, on the one hand, and intuitions that support the Bodily Continuity Theory, on the other.¹⁵ Thus, the case has been used to argue that our intuitions about personal identity puzzle cases are unreliable (Rovane, 1998; Szabó Gendler, 1998), or that it is indeterminate which of the Psychological and Bodily Continuity Theories is true (Sider, 2001a). But as our third presentation shows, Williams's puzzle arises not because we're torn between the Psychological and Bodily Continuity Theories, but because we're torn between the Complex View (in particular, the Psychological Continuity Theory) and the Simple View.

We ought to have been suspicious of Williams's diagnosis of his puzzle anyway, since it simply says nothing about *why* the psychological continuity intuitions go with the third-personal presentation of the case, or why the bodily continuity intuitions go with the first-personal presentation of the case. This, after all, is one of the most striking features of Williams's puzzle. The 'first-person/third-person' contrast will be integral to our explanation

¹⁵ Madell (1981) is a notable exception.

of this puzzle.

4.4 Imagination and possibility

How should we begin our search for a solution to this puzzle? Note that the conflict is over the number of possibilities for the person facts that are compatible with a given specification of the continuity facts. When I imagine these cases from the outside, I see only one ‘person possibility’ consistent with the specified continuity facts, but when I imagine these cases from the inside, it seems that there are multiple possibilities for my survival consistent with the specified continuity facts.

Given the abstract structure of the puzzle, what might a solution to it look like? Here is a suggestion: Maybe these two types of imagining correspond to two different kinds of metaphysical possibility. Perhaps while only one ‘person possibility₁’ is compatible with the specified continuity facts in a given case, several ‘possibilities₂’ for my persistence are compatible with those same continuity facts. And maybe when we look at a case from the outside we see the possibilities₁, and when we look from the inside we see the possibilities₂. Something like that, we might agree, would solve the puzzle if it were true. But why should we think any like this is true? What would these two types of possibility even be?

I think that a plausible answer to that question can be had if we consider the connection between imagination and possibility in light of the account of the distinction between imagining from the inside and imagining from the outside that we developed in §§1.1-1.2. Recall that account:

Uncentered Imagination

The content of an imagining from the outside is a set of possible worlds.

For all imaginings I , I is an imagining from the outside iff I ’s content is a set of possible worlds.

Centered Imagination

The content of an imagining from the inside is a set of centered worlds.

For all imaginings I , I is an imagining from the inside iff I ’s content is a set of centered worlds.

Before I go on to say something about the connection between this account of imagination and the nature of possibility, I want to say the following

about this account. In Chapter 1, I noted that there was a problem for this account, the problem of the impersonal-*de se*. And in Chapters 2 and 3, I offered two different solutions to this problem, but the resulting accounts were more complex than the account repeated above. But the problematic cases that motivated those more complex accounts only really arise when one considers how the inside-outside distinction interacts with *de se* ignorance. As I said in §1.2.3, in cases where such ignorance is not at issue, this simple account, **Uncentered and Centered Imagination**, seems entirely adequate (compare the fact that it seems fine to take all contents to be possible worlds propositions when one is not theorizing about the *de se*). Since the issues we are investigating here have little to do with *de se* ignorance, I think we can ignore some of the complexities discussed in the earlier chapters, and employ this simple view of the inside-outside distinction.

On to possibility. Now while there is much discussion in the philosophical literature about the relationship between imaginability and possibility,¹⁶ there is virtually none about what metaphysical consequences (if any) the inside-outside distinction has. But given our account of this distinction, an intriguing idea is that each type of imagining is a guide to a different kind of metaphysical possibility. Perhaps just as there are centered and uncentered *contents*, there are centered and uncentered *metaphysical possibilities*.

Suppose that, normally, imagining *p* provides evidence that *p* is possible, no matter what kind of content *p* is. When I imagine *p* from the third-person perspective, *p* is a *de dicto* content, something that has a truth value at a possible world. So one might take this imagining as evidence that there is a *possible world* accessible from the actual world at which *p* is true. Here, *p* is an uncentered possibility, a possible way for *the world* to be. But when I imagine *p* from the first-person perspective, *p* is a *de se* content, something that only has a truth value at a centered world. So one might take this imagining as evidence that there is a *centered world* accessible from ⟨actual world, me⟩ at which *p* is true.¹⁷ Here, *p* is a centered possibility, a possible way for *me* to be. From the outside, I see the ways the world could be; from the inside, I see the ways I could be.

Following Lewis (1983a, 1986, §4.4), let us suppose that uncentered and

¹⁶For some recent discussion, see the articles in Szabó Gendler and Hawthorne (2002).

¹⁷The accessibility relation is just a place-holder at the moment; we'll discuss what it means shortly.

4. Imagination, Possibility, and Personal Identity

centered possibilities are both genuine species of metaphysical possibility.¹⁸ Here's the basic idea:

Uncentered Possibility

Some possibility claims can be analyzed in terms of quantification over possible worlds.

Centered Possibility

Some possibility claims can be analyzed in terms of quantification over centered worlds.

Now this distinction is only interesting if these two types of possibility can come apart, if some claims of the form *I could have been F* are true in one sense but not the other. One type of a possibility claim that *could* be analyzed as a true centered possibility claim, but which would be false when analyzed as an uncentered possibility claim, would be a claim like *I could have been Fred*, as uttered by someone other than Fred. Here's how we would analyze that claim as a centered possibility claim:

I could have been Fred is true as uttered by x_c in w_c iff there is a centered world $\langle w, x \rangle$ accessible from $\langle w_c, x_c \rangle$ such that x is Fred in w .¹⁹

On the standard possible worlds analysis, that claim would be false when uttered by anyone other than Fred. Given that I am not Fred, and given the necessity of distinctness, there is no possible world in which I am Fred, and so, in the uncentered sense, I could not have been Fred.

¹⁸See also Hazen (1979). The Hazen-Lewis proposal is conducted within counterpart theory. Our account is the centered worlds ‘translation’ of theirs, just as our account of the *de se-de dicto* distinction is the centered worlds translation of Lewis (1979). The centered worlds framework does not require counterpart theory’s metaphysic of world-bound individuals (cf. §3.1). But on at least one reasonable understanding of the term *counterpart theory* (Fara, 2007), our centered possibility proposal counts as a kind of counterpart theory.

¹⁹Note that nothing in our apparatus demands that we think of centered possibility as an essentially *first-personal* kind of metaphysical possibility. The above analysis of centered possibility does not depend on the modal claim’s being expressed with the first-person singular pronoun *I*. This can be seen by looking at how we could analyze *Dilip could have been Fred* as a centered possibility claim:

Dilip could have been Fred is true at w_c iff there is a centered world $\langle w, x \rangle$ accessible from $\langle w_c, Dilip \rangle$ such that x is Fred in w .

See the Appendix for more on the semantics of centered possibility.

But as several philosophers have noted, the idea that (as each of us would put it) I could have been someone else is not without intuitive support:

...my being TN (or whoever in fact I am) seems accidental... So far as what I am essentially is concerned, it seems as if I just *happen* to be the publicly identifiable person TN—as if what I really am, this conscious subject, might just as well view the world from the perspective of a different person. (Nagel, 1986, 60)

Here am I, there goes poor Fred; there but for the grace of God go I; how lucky I am to be me, not him. Where there is luck there must be contingency. I am contemplating the possibility of my being poor Fred, and rejoicing that it is unrealized... (Lewis, 1983a, 395)

‘I might have been somebody else’ is a very primitive and very real thought; and it tends to carry with it an idea that one knows what it would be like for this ‘I’ to look out on a different world, from a different body, and still be the same ‘I’. (Williams, 1966, 40)

Now you don’t *have to* accept this idea in order to accept Centered Possibility. But if you’re inclined to accept it – or to at least think it’s coherent – then our account gives you a way of making sense of this thought.

We arrived at this ‘two kinds of possibility’ proposal by reflecting on the connection between imagination and possibility. We can now set out this connection as follows:

Uncentered Guide

Imagining from the outside is a guide to uncentered possibility. If I can imagine an uncentered content p , that is evidence that there is a possible world w accessible from the actual world such that p is true at w .²⁰

²⁰Many philosophers believe: (i) that every possible world is metaphysically accessible from every other possible world, and (ii) that ‘possible world’ and ‘metaphysically possible world’ are co-extensive. As a result, these philosophers do not need to mention

Centered Guide

Imagining from the inside is a guide to centered possibility. If I can imagine a centered content p , that is evidence that there is a centered world $\langle w, x \rangle$ accessible from $\langle \text{actual world}, \text{me} \rangle$ such that p is true at $\langle w, x \rangle$.

One worry about our analysis is that it does not explain the notion of centered possibility in independently understood terms. What is the relevant accessibility relation here? Metaphysical accessibility, which relates a possible individual to each of the possible individuals she could have been? If that is the only answer we can give, as I suspect it is, then the analysis is to some degree circular. But in this, the centered worlds analysis is not much different from the possible worlds analysis. Possible worlds semantics analyzes:

It is possible that Aristotle could have died as a child

as:

There is a possible world (accessible from the actual world) in which Aristotle died as a child.

But the latter obviously does not explain the former in independently understood terms. In both cases – centered and uncentered – the ‘analyses’ only serve to clarify the structure of the relevant modal claims.²¹

A different sort of question about the accessibility relation concerns not the *analysis* of the accessibility relation, but its *extension*: which centered worlds are accessible from me? Could I have been an aggregate of person stages that is neither psychologically nor bodily continuous? Could I have been Napoleon? Could I have been a poached egg?

I do not have any principled way of determining the extension of the accessibility relation. But I don’t think that our entitlement to appeal to the notion of centered possibility stands or falls with our ability to precisely

accessibility relations in their analyses of modal claims. But some theorists reject (i) and/or (ii), in which case the reference to a metaphysical accessibility relation is not redundant. See, for example, Salmon (1998) and the references therein. Since both views can be represented in terms of a metaphysical accessibility relation, I formulate uncentered possibility claims in those terms.

²¹At any rate, this is the line most actualists take. Lewis, on the other hand, aims to reduce the modal to the non-modal, and so would reject this as a characterization of his possible worlds analysis of modality. I side with the actualists here.

delineate the extension of the accessibility relation. Again, it is instructive to compare this situation with the case of possible worlds. Few philosophers would think they know exactly what the space of possible worlds is like. Are there possible worlds where Adam has all the qualitative properties Noah has, and vice-versa? Are there possible worlds physically identical to this one, but in which no one is conscious? Are there possible worlds almost exactly like this one but in which it rains beer throughout the earth for fifteen minutes in 1976?

I do not know the answers to these questions. For some of them, I do not even know how to go about answering them. Despite that, I think I do know certain possible worlds claims, e.g. that there is a possible world in which my sister is a school teacher, that there are no possible worlds in which Barack Obama is a non-human robot, and that no two possible worlds have exactly the same physical facts but different moral facts (the moral supervenes on the physical). The legitimacy of using the possible worlds apparatus to elucidate *particular* modal claims which one accepts doesn't require one to answer every question of the sort "But is there a possible world in which such-and-such happens?" Similarly, unless one has an antecedent reason to be skeptical of the notion of centered possibility, it seems legitimate to use the centered possibility apparatus to elucidate particular centered possibility claims even if one cannot answer every question of the sort "Is there a centered world accessible from you centered on a being like *this*?"

But let me say this: if we think of imagining from the inside as our basic way of representing centered possibilities, then it is natural to assume that accessibility is a relation between centered worlds that are centered on things that have *perspectives* or *points of view*.²² This might be a necessary condition on the accessibility relation: any centered world $\langle w, x \rangle$ accessible from me must be such that x has a 'perspective' in w (whatever that amounts to). Whether this is also a sufficient condition is a question we can leave unanswered for present purposes. This condition is, of course, seriously under-specified insofar as we lack an account of what it takes to have a 'perspective'. But, at the very least, this view would seem rule out the possibility that there are centered worlds accessible from me that are centered on poached eggs or bedposts. This view of accessibility thus differs from Lewis's construal, according to which every possible object is accessible from every other (Lewis, 1986, 239 - 243). I think there are tradeoffs between these two views: Lewis's has the advantage of not having

²²Thanks to Seth Yalcin for discussion on this matter.

to explain a restriction, but the disadvantage of alienating the proposal from the first-personal intuitions that motivate it in the first place.

Two other possible constraints on accessibility that are worth considering are these:

- *Origins*: any centered world $\langle w, x \rangle$ accessible from me must be such that x 's origins in w are sufficiently similar to mine.
- *Continuity*: any centered world $\langle w, x \rangle$ accessible from me must be such that x in w is an aggregate of person stages whose stages are maximally interrelated by some relatively ‘natural’ continuity relation (e.g. bodily continuity, branching psychological continuity, non-branching psychological continuity, etc.).

It's worth noting that the solution to the puzzle about *Supervenience* that I will present shortly will go through even if we adopted these two constraints along with the perspective constraint. This would leave us with a fairly conservative notion of centered possibility, one which would not countenance the *I could have been Fred* intuition.

There may be other worries about the idea of centered possibility, worries which are independent of questions about the analysis and extension of the accessibility relation. But I think it is legitimate to postpone questions about the ‘ultimate tenability’ of this proposal, at least for the moment. I think this is legitimate because whether or not we should adopt *Centered Possibility* depends in part on whether doing so would make other philosophical problems more tractable. And in the remainder of the paper I shall argue that this hypothesis helps us solve at least one problem—our puzzle about *Supervenience*.²³

²³Another question that sometimes comes up is whether it would be better to understand centered possibilities as *conceptual*, rather than *metaphysical*, possibilities. This is a much bigger question that I can hope to deal with here, not least because of the complexities involved in drawing the conceptual-metaphysical distinction. But here's how I view the present project: We come across some possibility judgments whose joint truth seems to conflict with *Supervenience* (e.g. *I could have undergone fission and survived as Lefty* and *I could have undergone fission and survived as Righty*). We arrive at those judgments by imagining the relevant hypothetical cases from the inside. Since imagining from the inside is imagining a centered content, it is plausible to suppose that the relevant possibility judgments should also be analyzed in terms of centered worlds. As I will demonstrate in §4.5, an important benefit of this analysis is that it renders the joint truth of those judgments compatible with *Supervenience*.

If this same story can be told while understanding the relevant possibility claims as conceptual possibility claims, then that would be acceptable to me, though I would like

4.5 The solution

Suppose we accept the foregoing claims about imagination and possibility. How does this help solve our puzzle? It turns out that this picture alone won't solve the puzzle—we need one more claim about what it is we imagine being when we imagine these cases from the inside.

4.5.1 One more claim

In §3.2, I noted that some philosophers think that when I imagine from the inside begin Napoleon, I am not imagining anything about the person I in fact am, Dilip. I am not imagining the truth of some actually false identity statement. I repeat the passage from Velleman that I quoted earlier:

If my approach to imagining that I am Napoleon... is to imagine *being* Napoleon, then I simply imagine a particular experience as experienced *by* Napoleon. I imagine the landscape as seen through Napoleon's eyes, the sounds of battle as heard through his ears, the nap of a tunic as felt by his hand. Although Napoleon doesn't appear in the resulting mental image, he does appear in the content of my imagining, since I am imagining Austerlitz as experienced by him. But I, David Velleman, am absent both from the image and from the content of the imagining: I'm not imagining anything about the person who I actually am. (Velleman, 1996, 40)

Velleman also connects this point to first-personal personal identity thought experiments of the sort we've been discussing, and it is this connection that will be crucial for us:

If I can imagine that I am Napoleon without imagining a Napoleonic identity for myself, then maybe I can anticipate that I will wake up in the future without anticipating a future for my actual

to learn more about how to draw the conceptual-metaphysical distinction, and why we should think that centered possibility claims always fall on the conceptual side of the line. Presumably, we are not always required to offer non-question-begging answers to questions of the form “how do you know that’s a metaphysical, rather than a conceptual, possibility?”. For if we are always subject to that requirement, I can’t see how skepticism about what’s metaphysically possible is to be avoided. For how do you defend, e.g., the claim that *Humprey could have won the election* is a true metaphysical possibility claim, and not merely a true conceptual possibility claim?

self, either... the thought's being first-personal doesn't guarantee that it is about me, the thinker... Perhaps the anticipation that I will wake up in the future can be similarly first-personal about a future subject who may or may not be identical with me. (Velleman, 1996, 41)

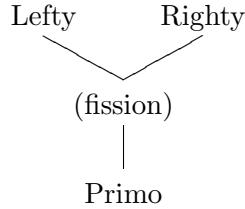
To connect Velleman's point with our discussion, consider the first-personally presented fission case. Suppose I imagine being the initial pre-fission person, and then waking up as one of the post-fission 'branches'. I take it that Velleman's insight, when applied to this case, is that when I imagine waking up after the fission procedure has been carried out, there is, as it were, no guarantee that the person I imagined being prior to the experiment is the same person as the person I've imagined being after the experiment, i.e. there is no guarantee that these are two stages of the same person. For if I can imagine being someone else, then surely I can imagine being the initial pre-fission person and then imagine being one of the post-fission branches, even if those two person stages are not two stages of a single person.

But one is inclined to object: when you imagine the first-personally described fission case, you don't *first* imagine being the initial pre-fission person and *then*, in a second act of imagining, imagine being one of the post-fission branches; rather, you simply imagine, in a single imaginative act, being the initial person and then waking up as one of the branches. You imagine *surviving as* one of the branches. I agree with the objection, but I think the basic point still stands. There is not, I think, a strict argument from the fact that I can imagine being someone else to the claim that I can imagine being one person at one time and then being another at a later time. But the fact that I can do the former *suggests* that I should be able to do the latter: it's plausible to think that the two abilities go together.

4.5.2 Fission

How does Velleman's point help with our puzzle? Consider this diagrammatic presentation of the fission case, where Primo is the stage of the initial person who undergoes fission, and Lefty and Righty are the two post-fission person stages that result.²⁴

²⁴Like *Lefty* and *Righty*, *Primo* is not a name but an abbreviation for a definite description: *the initial person who undergoes fission*.



When we imagine the fission case impersonally, we judge that neither Primo and Lefty nor Primo and Righty are stages of the same person. But when we imagine the case from the first-person perspective, we imagine, for example, first being Primo and then being Lefty, even though Primo and Lefty are distinct persons.

On this interpretation, then, the fission diagram represents the *objective scenario* that one imagines when one imagines the case from the third-person perspective. And it is this *same scenario* one imagines when one imagines the case from the first-person perspective. When one looks at the case from the first-person perspective, one can imagine traveling along one of the ‘paths’ represented in the diagram, either from Primo to Lefty or from Primo to Righty. One can imagine traveling along as a space-time worm composed of two person stages which do not constitute a person. But the path one travels along is genuinely contained in the same objective scenario that one imagines when one imagines the case from the third-person perspective.

Given a four-dimensionalist metaphysic of persistence, there is an aggregate object composed of Primo and Lefty and an aggregate object composed of Primo and Righty.²⁵ Call these aggregates *Primo + Lefty* and *Primo + Righty* respectively. Given the existence of *Primo + Lefty*, it is tempting to suppose that when I imagine surviving fission as Lefty, I am imagining being *Primo + Lefty*—I am imagining being an aggregate of person stages which does not constitute a person. Similarly, when I imagine surviving as Righty, I’m imagining being *Primo + Righty*. Let’s set these claims out as follows:

- (13) a. When I imagine from the inside surviving fission as Lefty, I am imagining being *Primo + Lefty*.
- b. When I imagine from the inside surviving fission as Righty, I am imagining being *Primo + Righty*.

²⁵Not everyone will accept the existence of these aggregate objects. I discuss this matter in §4.5.4.

4. Imagination, Possibility, and Personal Identity

Now I think we have the materials to solve the puzzle about Supervenience, at least as it concerns the fission case.

The problem, recall, is this: when we look from the outside, Supervenience seems true; but when we look from the inside, Supervenience seems false. But let's look more closely at *why* Supervenience seems false when we look from the inside. It seems false because there seem to be multiple 'person possibilities' consistent with the same continuity facts. There are at least two possibilities:

- (14) a. I could undergo fission and survive as Lefty.
- b. I could undergo fission and survive as Righty.

(There's also a third possibility: I undergo fission and don't survive. We leave this aside for the sake of brevity, since the above two possibilities are enough to generate an apparent problem for Supervenience.) If we make two assumptions, then the joint truth of (14a) and (14b) is indeed incompatible with Supervenience. The first assumption is that I am essentially a person—there is no possible world in which I am not a person. The second is that these possibility claims must be analyzed in terms of quantification over possible worlds.

Now consider what it would take for both (14a) and (14b) to be true, given those two assumptions: there would have to be possible worlds, w and w' , such that in w , I undergo fission and survive as Lefty, but in w' , I undergo fission and survive as Righty. So consider x , a pre-fission stage of me in w , and y , a post-fission stage of Lefty in w . Consider also x' , a pre-fission stage of me in w' , and y' , a post-fission stage of Lefty in w' . Note that x and y in w and x' and y' in w' are exactly the same with respect to continuity. Now note two things: First, x and y are related by the relation of personal identity in w , since I end up as Lefty in w , and (by assumption) I am a person in w . Second, x' and y' are *not* related by the relation of personal identity in w' , since I end up as Righty, not as Lefty, in w' , and (by assumption) I am a person in w' . Thus, we have an apparent counterexample to Supervenience.

It should be clear where this line of reasoning breaks down, according to our approach. The problem is with the assumption that these possibility claims should be analyzed in terms of quantification over possible worlds. Since these possibilities are seen from the inside, they are centered possibilities; claims (14a) and (14b) should thus be analyzed in terms of quantification over centered worlds. Given our claim that when I imagine surviving as Lefty, I'm imagining being Primo + Lefty, and that when

I imagine surviving as Righty, I'm imagining being Primo + Righty, we ought to analyze (14a) and (14b), respectively, as (15a) and (15b):

- (15) a. There is a centered world $\langle w, x \rangle$ accessible from $\langle \text{actual world}, \text{me} \rangle$ such that Primo undergoes fission in w , and x is Primo + Lefty in w .
- b. There is a centered world $\langle w, x \rangle$ accessible from $\langle \text{actual world}, \text{me} \rangle$ such that Primo undergoes fission in w , and x is Primo + Righty in w .

But this means that our judgments about the first-personally presented fission case pose no threat to **Supervenience**. For if (14a) and (14b) are to be analyzed, respectively, as (15a) and (15b), then clearly the joint truth of (14a) and (14b) does not require the existence of two pairs of possible person stages which are the same continuity-wise but which differ person-wise. Instead, the conjunction of those two claims requires only the existence of two accessible centered worlds, one of which is centered on something that ends up with the original left hemisphere, the other which is centered on something which ends up with original right hemisphere. And the existence of two such accessible centered worlds is perfectly compatible with the truth of **Supervenience**.

What about the first assumption mentioned above: that I am essentially a person? We glossed this as the claim that there is no possible world in which I am not a person. Given *that* gloss on *essentially*, then it is true to say that I am essentially a person—at least, nothing I want to say conflicts with that claim. But now that we have *two* notions of metaphysical possibility, there is another way of understanding the claim that I am essentially a person—namely, that no centered world accessible from me is centered on a non-person. And this, according to our approach, is false. Primo + Lefty, for example, is not a person, and yet I could have been him; thus, a centered world centered on that non-person is accessible from me. That some non-person aggregates of person stages are possible ways for me to be is the result of combining Velleman's observation – I can imagine being one person and then another – with the idea that imagining from the inside is a guide to centered possibility.

4.5.3 Williams's puzzle

A similar solution to Williams's puzzle can be given. Imagining Williams's case from the outside seems to support **Supervenience**, since only one per-

son possibility seems compatible with the continuity facts: the two initial persons switch bodies. But when I imagine this case from the inside, I can imagine being *preA* and ending up as either *postA* (second presentation) or *postB* (third presentation).²⁶ Thus, I see two possibilities for my survival, both of which are consistent with the same continuity facts. This leads me to endorse two possibility claims:

- (16) a. I could start out as *preA* in a Williams case, and end up as *postA*.
- b. I could start out as *preA* in a Williams case, and end up as *postB*.

(Again, there is a third possibility – I could be *preA* and not survive – which we ignore for the sake of ease.) Again, the problem arises against the background of two assumptions: that the possibility claims in (16) should be analyzed in terms of possible worlds, and that I am essentially a person.

Given those assumptions, the conjunction of (16a) and (16b) poses a threat to Supervenience. For the truth of (16a) requires there to be a world *w* in which my initial pre-procedure stage *preA* is person-related to *postA*, while the truth of (16b) will require there to be a world *w'* in which my initial pre-procedure stage *preA* is *not* person-related to *postA* (for in *w'*, I end up as *postB*). So even though *preA* and *postA* in *w* and *preA* and *postA* in *w'* are the same continuity-wise, they are not the same person-wise—in *w*, they are stages of the same person, whereas in *w'*, they are not. And this conflicts with Supervenience.

The resolution is as before—we reject the assumption that the possibility claims in (16) are uncentered possibility claims, since what we see from the inside are centered possibilities. And we assume that when I imagine ending up as *postA*, I am imagining being an aggregate of person stages that does not constitute a person, namely *preA + postA* (this is the Velleman point, applied to Williams's case). And when I imagine ending up as *postB*, I am imagining being *preA + postB*, an aggregate of person stages that happens to be a person.

Thus, (16a) and (16b) should be analyzed as (17a) and (17b) respectively:

- (17) a. There is a centered world $\langle w, x \rangle$ accessible from $\langle \text{actual world}, \text{me} \rangle$ such that *x* is *preA + postA* in a Williams case in *w*.

²⁶Recall our abbreviations: *preA* is the person in the *A*-body prior to procedure, *postA* the person in the *A*-body after the procedure; '*preB*' and '*postB*' are defined similarly. (Imagine again that the bodies have large scarlet letters painted on them.)

- b. There is a centered world $\langle w, x \rangle$ accessible from \langle actual world, me \rangle such that x is $preA + postB$ in a Williams case in w .

On this interpretation, the conjunction of (16a) and (16b) does not require the existence of two pairs of possible person stages which are the same continuity-wise, but differ person-wise. Again, the conjunction of those two claims requires only the existence of two accessible centered worlds, which are centered on different sorts of aggregates of person stages. One is centered on $preA + postA$, a ‘body’ aggregate (i.e. an aggregate of person stages that is maximally interrelated by the relation of bodily continuity) that is not a ‘psychological’ aggregate (i.e. an aggregate of person stages that is maximally interrelated by the relation of psychological continuity); the other is centered on $preA + postB$, a ‘psychological’ aggregate that is not a ‘body’ aggregate. But the existence of these two accessible centered worlds in no way threatens Supervenience.

4.5.4 Three-dimensionalism

That’s my solution to the puzzle. Before going on to discuss how this bears on the Simple-Complex dispute, I want to return to an issue I discussed briefly at the outset: the dispute between three-dimensionalism and four-dimensionalism.

Throughout this discussion, we have used the four-dimensionalist’s language of person stages and aggregates. Indeed, we framed the central debate as one over Supervenience, which we stated in the language of person stages. But, of course, three-dimensionalists will not accept this talk of person stages at face value. Fortunately, there is a way of interpreting Supervenience that should be acceptable to three-dimensionalists. Suppose we accept the idea that, for each person x , there is an event which is the *life history* of x (cf. Perry, 1975, Introduction). A life history for x can be thought of as a series of ‘total states’ of x , one state for each moment at which x exists. Then three-dimensionalists can understand a ‘person stage’ as a total state of a person at a time, and they can understand Supervenience not as a claim about persons but as a claim about life histories. On this interpretation, Supervenience says that fixing the continuity relations between a total person state x at time t and a total person state y at a distinct time t' suffices to determine whether x and y are two parts of the life history of a single person.²⁷

²⁷For more on how three-dimensionalists should understand supervenience theses concerning persistence, see Zimmerman (1998) and Sider (1999).

This shows we can frame *the debate* in terms acceptable to three-dimensionalists. But can three-dimensionalists accept our *solution* to the puzzle? Some three-dimensionalists certainly can: our solution should be acceptable to three-dimensionalists who endorse what Sider (2001a) calls *promiscuous endurantism*.²⁸ Promiscuous endurantism says that:

...in the vicinity of every person [there is] a plurality of coincident entities, which share the same momentary properties but differ in their persistence conditions. In my vicinity, there is a psychological-person, a body-person, and perhaps other entities corresponding to other criteria of personal identity. (Sider, 2001a)

To see that our solution is available to the promiscuous endurantist, let's look at our treatment of the fission case. One of the crucial moves we made was to say that, when I look at the fission case from the inside, I see multiple centered possibilities: one centered on Primo + Lefty and one centered on Primo + Righty. The four-dimensionalist understands these objects to be aggregates of persons. But the promiscuous endurantist can understand *Primo + Lefty* as follows: if Primo begins to exist at t_1 , fission occurs at t_2 , and Lefty ceases to exist as t_3 , then *Primo + Lefty* picks out an enduring object that is wholly spatially coincident with Primo from t_1 to t_2 , and then is wholly spatially coincident with Lefty from t_2 to t_3 , at which time it too ceases to exist. The promiscuous endurantist can say something similar about *Primo + Righty*. Note that these objects, Primo + Lefty and Primo + Righty, correspond to Lewis's (1976) *branching* psychological criterion of personal identity, and so the promiscuous endurantist will accept their existence.

But not every endurantist is promiscuous; some are *chaste*. Sider (2001a) defines *chaste endurantism* as three-dimensionalism plus the claim that person-like entities *never* wholly spatially coincide.²⁹ The chaste endurantist does not believe that there are any objects like Primo + Lefty or Primo + Righty, as understood by either the four-dimensionalist or the promiscuous endurantist. Thus, the chaste endurantist will not be able to help herself to centered worlds centered on Primo + Lefty and Primo +

²⁸Thanks to Caspar Hare for helping me to see this point.

²⁹Olson (1997) seems to endorse something like chaste endurantism. Promiscuity and chastity do not exhaust endurantist's options, of course. Whether our solution to the puzzle is compatible with a given intermediate endurantist position will depend on the precise details of that position.

Righty, for on her view there are no such centered worlds since the relevant ‘centers’ do not exist. So it seems that our solution – at least in its present formulation – is not available to the chaste endurantist.

So while our solution is independent of a specific view of *how things persist*, it is not independent of a specific view of *what there is*. We require a suitably generous ontology. That said, there *may* be a way to use our framework of centered/uncentered imagination and possibility to fashion a solution to the puzzle that would be acceptable to the chaste endurantist. But I leave that as a matter for future inquiry.³⁰

4.6 Something like the Simple View

Recall the Velleman point: when one imagines personal identity cases from the inside, there is no guarantee that the person one imagines starting out as at the beginning of the case is the same person as the person one imagines ending up as at the end of the case. What this suggests is that what I can imagine surviving from the inside is no guide to what sort of vicissitudes a *person* can survive. In other words, first-personal intuitions about what I could survive are not intuitions about personal identity at all; rather they are intuitions that concern possible ways for me to be.

If this is right, then it appears to vindicate the Complex View: since our third-person intuitions favor Supervenience, and since our first-person intuitions simply don’t bear on the truth of Supervenience, we ought to accept the Complex View. But even though we have to reject the letter of the Simple View, our account nevertheless vindicates the basic thought that motivates it. According to our account, it is true that, in a given possible situation, who I will be in the future is not settled by the continuity relations I bear to future person stages. So there is a sense in which something like the Simple View is true.

We characterized the Simple View as the denial of a supervenience thesis, so it would be nice to characterize our something-like-the-Simple-

³⁰One possibility is that we represent centered possibilities not using centered worlds, but by using pairs of a world w and a function f from times to individuals in w . Here $\langle w, f \rangle$ represents the possibility that, for each time t for which f is defined, I am $f(t)$ at t in w . So if it’s possible for me to be Primo at t_1 , and then Lefty at t_2 , then there is a $\langle w, f \rangle$ accessible from me where w is a world in which fission occurs, and f is the function $\{\langle t_1, \text{Primo} \rangle, \langle t_2, \text{Lefty} \rangle\}$. This proposal doesn’t require there to be objects like Primo + Lefty. But on this proposal, we cannot ask which *individual* I am in the centered possibility $\langle w, f \rangle$, but only which individual I am at such-and-such a time. Something like this might work, though one would need to think through the details carefully.

View-view in a similar fashion. But what would the relevant supervenience thesis be? Given that *supervenience* is a philosophical term of art, it's not clear that what we're going to formulate really counts as a supervenience thesis. Since the intuitive idea behind supervenience is that fixing one set of facts fixes another set of facts, let's call the relevant thesis a *fixity thesis*. Our fixity thesis should say something like this: Fix who I am at a given time t in a possible situation. That is, fix that, at t , I am person stage x_t (I am some aggregate x whose temporal part at t is x_t). Then fix all the continuity relations that x_t bears to every other person stage y in w . Then you've fixed whether or not y is a temporal part of x (i.e. me) or not. Here is the precise formulation:

Fixity

For all centered worlds $\langle w, t, x \rangle, \langle w', t', x' \rangle$ accessible from \langle actual world, now, me \rangle and person stages y in w and y' in w' : if x_t and y in w and $x'_{t'}$ and y' in w' are the same with respect to continuity, then y is a part of x in w just in case y' is part of x' in w' .

Note that here we use the ‘official’ notion of a centered world, a world-time-individual triple, rather than simply a world-individual pair (see footnote 3 in §1.1).

If Fixity were true, then once you fixed who I was in a given centered situation at a time t , and fixed the relations of continuity between me-at- t and every other person stage, then you would have determined who I would be in the future (and who I had been in the past). But Fixity is not true, at least not if we take our first-person intuitions at face value. For Fixity is undermined by the judgments we arrive at when we consider fission and Williams’s puzzle from the inside.

Take fission. I can imagine surviving fission as Lefty or as Righty. So there are two centered situations or *predicaments* I could be in, a Primo-Lefty predicament and a Primo-Righty predicament (a predicament is represented by a set of centered worlds). To get a counterexample to Fixity, consider two centered worlds accessible from me, $\langle w, t, x \rangle$ in the Primo-Lefty predicament, $\langle w', t', x' \rangle$ in the Primo-Righty predicament. Take x_t to be Primo in w , and $x'_{t'}$ to be Primo in w' . And take y to be Lefty in w , and y' to be Lefty in w' . x_t and y in w and $x'_{t'}$ and y' in w' are the same with respect to continuity. In the Primo-Lefty predicament, I end up as Lefty. So y , i.e. Lefty, is a part of x in w , since x is Primo + Lefty in

w. But in the Primo-Righty predicament, I end up as Righty. So *y'*, i.e. Lefty, is *not* a part of *x'* in *w'*, since *x'* is Primo + Righty in *w'*.

Take the second and third presentations of Williams's case. In the second presentation, I start out as *preA* and end up as *postA*, while in the third presentation, I start out as *preA* and end up as *postB*. We get another counterexample to Fixity by considering two centered worlds accessible from me, $\langle w, t, x \rangle$ in the first predicament, $\langle w', t', x' \rangle$ in the second. Take x_t to be *preA* in *w*, and $x'_{t'}$ to be *preA* in *w'*. And take *y* to be *postA* in *w*, and *y'* to be *postA* in *w'*. Again, x_t and *y* in *w* and $x'_{t'}$ and *y'* in *w'* are the same with respect to continuity. But *y*, i.e. *postA*, is part of *x* in *w*, since *x* is *preA + postA* in *w*. But *y'*, i.e. *postA*, is not part of *x'* in *w'*, since *x'* is *preA + postB* in *w'*.

Does personal identity supervene on continuity? Yes, for Supervenience is true. Is the question of who I will be, or who I have been, settled by which continuity relations I bear to other person stages? No, for Fixity is false. So the Simple View is wrong, but it gets something right. As friends of the Simple View have insisted, there is an important sense in which who I will be, or who I have been, is not settled by what continuity relations I bear to other person stages. But when properly understood this thought supports the denial of Fixity rather than the denial of Supervenience. And denying Fixity, it should be clear, is compatible with accepting the Complex View, since Fixity simply says nothing about what it takes for two person stages to be stages of the same person. Once we are clear about the nature of first-personal intuitions about persistence, we see that the 'simplicity of the self' is compatible with the complexity of persons.

Appendix A

Semantics for centered possibility

This appendix sketches a semantics for the sort of centered possibility claims discussed in Chapter 4. Consider the following centered possibility claims:

- (18) a. I could have been Fred.
b. Dilip could have been Fred.

We want a theory that will yield the following truth conditions:

- (19) a. $\llbracket \text{I could have been Fred} \rrbracket^{c,i,g} = 1$ iff there is a $\langle w, x \rangle$ accessible from $\langle w_i, x_c \rangle$ such that x is Fred in w .
b. $\llbracket \text{Dilip could have been Fred} \rrbracket^{c,i,g} = 1$ iff there is a $\langle w, x \rangle$ accessible from $\langle w_i, \text{Dilip} \rangle$ such that x is Fred in w .

What would the semantics for the modal need to look like in order for us to derive (19a) and (19b)? I will focus on (19b). On the standard account of the semantics of modals found in Kratzer (1981, 1991), modals take as an argument a contextually supplied set of possible worlds called *the modal base*. The modal then quantifies over the worlds in that base, saying that the embedded sentence is true at one world in the base (if the modal is a possibility modal) or that it is true at every world in the base (if the modal is a necessity modal). One way to implement Kratzer's proposal is to assume that a modal always combines with a covert variable which is assigned a set of worlds by the variable assignment. So a schematic Kratzer-inspired entry for such a covert variable might look like this:

$$\llbracket B_1 \rrbracket^{c,i,g} = g(B_1) = \{w : w \text{ is accessible from } w_i\}$$

Since we now require modals to quantify over centered worlds, we can re-construe modal bases as sets of centered worlds. What we want is for the modal in (19b) to quantify over the following set of centered worlds:

$$\{\langle w, x \rangle : \langle w, x \rangle \text{ is accessible from } \langle w_i, \text{Dilip} \rangle\}$$

One way to get this result is to make $\llbracket B_1 \rrbracket^{c,i,g}$ not a set of centered worlds, but rather a *function* from an individual to a set of centered worlds: $\llbracket B_1 \rrbracket^{c,i,g}$ would take an individual y to the set centered worlds accessible from $\langle w_i, y \rangle$. We could then feed that function the individual denoted by *Dilip*. The possibility modal would then quantify over the centered worlds accessible from $\langle w_i, \text{Dilip} \rangle$, and say that at least one of those centered worlds $\langle w, x \rangle$ is such that x is Fred in w , which is exactly what we want.

The suggestion requires making sure that, in the LF of (18b), $\llbracket B_1 \rrbracket^{c,i,g}$ takes Dilip as an argument. This can be accomplished using the same sort of *res* movement posited by the theories of *de re* attitude ascription discussed in §2.6. When ‘Dilip’ moves, it leaves behind a trace, ‘ t_2 ’, which is bound by a λ -binder occurring at the top of the complement clause. Here is the lexical entry for ‘ B_1 ’ on this account, followed by a specification of the proposed structure for (18b):

$$\llbracket B_1 \rrbracket^{c,i,g} = \lambda y_e. \{ \langle w, x \rangle : \langle w, x \rangle \text{ is accessible from } \langle w_i, y \rangle \}$$

$$\text{Possibly } [B_1 \text{ Dilip }] [\lambda_2 t_2 \text{ is Fred}]$$

Given this structure and the assumed lexical entry, we get the following truth condition:

$$\llbracket \text{Possibly} \rrbracket^{c,i,g} (\llbracket B_1 \rrbracket^{c,i,g} (\llbracket \text{Dilip} \rrbracket^{c,i,g})) (\lambda w. \llbracket \lambda_2 t_2 \text{ is Fred} \rrbracket^{c,w,g}) = 1 \text{ iff}$$

$$\llbracket \text{Possibly} \rrbracket^{c,i,g} (\{ \langle w, x \rangle : \langle w, x \rangle \text{ is accessible from } \langle w_i, \text{Dilip} \rangle \}) (\lambda w. \llbracket \lambda_2 t_2 \text{ is Fred} \rrbracket^{c,w,g}) = 1 \text{ iff}$$

There is a $\langle w', x' \rangle$ accessible from $\langle w_i, \text{Dilip} \rangle$ such that $[\lambda w. \lambda x. x \text{ is Fred}](w')(x') = 1$ iff

There is a $\langle w', x' \rangle$ that is accessible from $\langle w_i, \text{Dilip} \rangle$ such that x' is Fred in w' .

This is the desired result.¹

¹This semantics might also be useful for making sense of *de re* epistemic modal claims like *I might be Rudolph Lingens* and *That guy might be Ortcutt*.

Bibliography

- Abusch, D. (1997). Sequence of tense and temporal *de re*. *Linguistics and Philosophy*, **20**(1), 1 – 50.
- Anand, P. (2006). *De De Se*. Ph.D. thesis, MIT.
- Anand, P. and Nevins, A. (2004). Shifty operators in changing contexts. In R. B. Young, editor, *Proceedings from Semantics and Linguistic Theory XIV*. CLC Publications, Cornell University.
- Bach, K. (1997). Do belief reports report beliefs? *Pacific Philosophical Quarterly*, **78**, 215 – 241.
- Baker, L. R. (2000). *Persons and Bodies*. Cambridge University Press, New York.
- Blackburn, S. (1997). Has Kant refuted Parfit? In J. Dancy, editor, *Reading Parfit*. Blackwell, Oxford.
- Brueckner, A. (2005). Branching and the psychological approach to personal identity. *Analysis*, **65**(4), 294–301.
- Burge, T. (1977). Belief *de re*. *The Journal of Philosophy*, **74**(6), 338 – 362.
- Chalmers, D. (2002a). The components of content. In D. Chalmers, editor, *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press, New York.
- Chalmers, D. (2002b). On sense and intension. *Philosophical Perspectives*, **16**.
- Chierchia, G. (1989). Anaphora and attitudes *de se*. In R. Bartsch, J. van Benthem, and van Emde Boas, editors, *Semantics and Contextual Expression*. Foris, Dordrecht, Netherlands.
- Chisholm, R. (1969). The loose and popular and the strict and philosophical senses of identity. In N. Care and R. Grimm, editors, *Perception and Personal Identity*, pages 82 – 106. The Press of Case Western Reserve University, Cleveland, OH.
- Chisholm, R. (1976). *Person and Object*. Open Court, La Salle, IL.
- Chisholm, R. (1981). *The First Person*. The Harvester Press, Brighton.
- Cresswell, M. J. and von Stechow, A. (1982). *De re* belief generalized. *Linguistics and Philosophy*, **5**(4), 503 – 535.
- Crimmins, M. (1992). *Talk About Beliefs*. MIT Press, Cambridge, MA.

- Donnellan, K. S. (1966). Reference and definite descriptions. In Ludlow (1997), pages 281 – 304.
- Egan, A. (2006). Secondary qualities and self-location. *Philosophy and Phenomenological Research*, **72**(1), 97 – 119.
- Fara, D. G. (2007). Semantics for context-dependent relative-sameness counterpart theory. Unpublished manuscript, Princeton University.
- von Fintel, K. (2005). Lecture notes for pragmatics in linguistic theory. Notes for class taught at the LSA Summer Institute, MIT.
- von Fintel, K. and Heim, I. (2004). Intensional semantics lecture notes. Notes for class taught at MIT.
- Hazen, A. (1979). Counterpart-theoretic semantics for modal logic. *The Journal of Philosophy*, **76**(6), 319 – 338.
- Heim, I. (1994a). Comments on Abusch’s theory of tense. In H. Kamp, editor, *Ellipsis, Tense, and Questions*. University of Amsterdam.
- Heim, I. (1994b). Puzzling reflexive pronouns in *de se* reports. Handout for Bielefeld conference, March 1994.
- Heim, I. and Kratzer, A. (1998). *Semantics in Generative Grammar*. Blackwell Publishers, Oxford, UK and Malden, MA.
- Higginbotham, J. (2003). Remembering, imagining, and the first person. In A. Barber, editor, *Epistemology of Language*, pages 496 – 533. Oxford University Press, Oxford and New York.
- Hill, C. (1997). Imaginability, conceivability, possibility and the mind-body problem. *Philosophical Studies*, **87**, 61 – 85.
- Hintikka, J. (1969). On the logic of perception. In *Models for Modalities*. Reidel, Dordrecht.
- Hintikka, J. (1975). *The Intentions of Intentionality and Other New Models for Modalities*. Reidel, Dordrecht.
- Kaplan, D. (1968). Quantifying in. *Synthese*, **19**(1-2), 178 – 214.
- King, J. C. (Fall 2006). Structured propositions. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*.
- Kratzer, A. (1981). The notional category of modality. In H.-J. Eikmeyer and H. Rieser, editors, *Words, Worlds, and Contexts*, pages 38–74. Walter de Gruyter and Co., Berlin and New York.
- Kratzer, A. (1991). Modality. In A. von Stechow and D. Wunderlich, editors, *Semantics: An International Handbook of Contemporary Research*. Walter de Gruyter, Berlin.
- Kripke, S. (1979). A puzzle about belief. In A. Margalit, editor, *Meaning and Use*. Reidel, Dordrecht.

Bibliography

- Lakoff, G. (1972). Linguistics and natural logic. In D. Davidson and G. Harman, editors, *Semantics of Natural Language*. D. Reidel Publishing Company, Dordrecht.
- Lewis, D. K. (1976). Survival and identity. In Lewis (1983b), pages 55 – 77.
- Lewis, D. K. (1979). Attitudes *de dicto* and *de se*. In Lewis (1983b), pages 133 – 159.
- Lewis, D. K. (1981). What puzzling Pierre does not believe. In Lewis (1999), pages 283 – 289.
- Lewis, D. K. (1983a). Individuation by acquaintance and by stipulation. In Lewis (1999), pages 373 – 402.
- Lewis, D. K. (1983b). *Philosophical Papers, Volume I*. Oxford University Press, New York and Oxford.
- Lewis, D. K. (1986). *On the Plurality of Worlds*. Basil Blackwell, Oxford.
- Lewis, D. K. (1994). Reduction of mind. In Lewis (1999), pages 291 – 324.
- Lewis, D. K. (1999). *Papers in Epistemology and Metaphysics*. Cambridge University Press, Cambridge, UK and New York.
- Ludlow, P., editor (1997). *Readings in the Philosophy of Language*. MIT Press, Cambridge, MA and London.
- Madell, G. (1981). *The Identity of the Self*. Edinburgh University Press, Edinburgh.
- Maier, E. (2006). *Belief in Context: Towards a Unified Semantics of De Re and De Se Attitude Reports*. Ph.D. thesis, Radboud University Nijmegen.
- Martin, M. (2002). The transparency of experience. *Mind and Language*, **17**(4), 376 – 425.
- McMichael, A. (1983). A problem for actualism about possible worlds. *The Philosophical Review*, **92**(1), 49 – 66.
- Morgan, J. (1970). On the criterion of NP deletion. *CLS*, **6**.
- Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, **83**, 435 – 450.
- Nagel, T. (1986). *The View from Nowhere*. Oxford University Press, New York and Oxford.
- Nolan, D. (2006). Selfless desires. *Philosophy and Phenomenological Research*, **73**(3), 665 – 679.
- Noonan, H. (1989). *Personal Identity*. Routledge, London.
- Olson, E. T. (1997). *The Human Animal*. Oxford University Press, New York and Oxford.
- Parfit, D. (1984). *Reasons and Persons*. Clarendon Press, Oxford.

- Peacocke, C. (1985). Imagination, experience, and possibility: A Berkeleyan view defended. In J. Foster and H. Robinson, editors, *Essays on Berkeley*, pages 19 – 35. Clarendon Press, Oxford.
- Perry, J., editor (1975). *Personal Identity*. University of California Press, Berkeley.
- Perry, J. (1977). Frege on demonstratives. In Perry (1993), pages 3 – 26.
- Perry, J. (1979). The problem of the essential indexical. In Perry (1993), pages 3 – 21.
- Perry, J. (1993). *The Problem of the Essential Indexical and Other Essays*. Oxford University Press, New York and Oxford.
- Pryor, J. (2004). An epistemic theory of acquaintance. Handout for a conference at UCSB.
- Quine, W. (1956). Quantifiers and propositional attitudes. In Quine (1976).
- Quine, W. (1969). Propositional objects. In *Ontological Relativity and Other Essays*. Columbia University Press, New York.
- Quine, W. (1976). *The Ways of Paradox and Other Essays*. Harvard University Press, Cambridge, MA.
- Recanati, F. (2007a). Imagining *de se*. Unpublished manuscript.
- Recanati, F. (2007b). *Perspectival Thought: A Plea for (Moderate) Relativism*. Oxford University Press.
- Reynolds, S. (1989). Imagining oneself to be another. *Noûs*, **23**(5), 615–33.
- Rovane, C. (1998). *The Bounds of Agency*. Princeton University Press, Princeton, NJ.
- Salmon, N. (1986). *Frege's Puzzle*. Ridgeview Publishing Company, Atascadero, CA.
- Salmon, N. (1998). The logic of what might have been. *The Philosophical Review*, **98**(1), 3 – 34.
- Schiffer, S. (1977). Naming and knowing. *Midwest Studies in Philosophy*, **2**, 28 – 41.
- Schlenker, P. (2003). A plea for monsters. *Linguistics and Philosophy*, **26**, 29–120.
- Schlenker, P. (2004). Person and binding (a partial survey). *Italian Journal of Linguistics/Rivista di Linguistica*, **16**(1).
- Shier, D. (1996). Direct reference for the narrow minded. *Pacific Philosophical Quarterly*, **77**, 225 – 248.
- Shoemaker, S. (1963). *Self-Knowledge and Self-Identity*. Cornell University Press, Ithaca, New York.
- Shoemaker, S. (1994). The first-person perspective. In Shoemaker (1996), pages 7 – 22.

Bibliography

- Shoemaker, S. (1996). *The First-Person Perspective and Other Essays*. Cambridge University Press, New York.
- Sider, T. (1999). Global supervenience and identity across time and worlds. *Philosophy and Phenomenological Research*, **59**, 913 – 937.
- Sider, T. (2001a). Criteria of personal identity and the limits of conceptual analysis. In J. Tomberlin, editor, *Philosophical Perspectives 15, Metaphysics*. Blackwell, Cambridge, MA.
- Sider, T. (2001b). *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford University Press, Oxford and New York.
- Sosa, E. (1970). Propositional attitudes *de dicto* and *de re*. *The Journal of Philosophy*, **67**(883 - 896).
- Stalnaker, R. (1988). Belief attribution and context. In R. Grimm and D. Merrill, editors, *Contents of Thought*. University of Arizona Press, Tucson.
- Swinburne, R. (1984). Personal identity: The dualist theory. In R. Swinburne and S. Shoemaker, editors, *Personal Identity*, pages 1–66. Basil Blackwell, Oxford.
- Szabó Gendler, T. (1998). Exceptional persons: On the limits of imaginary cases. *Journal of Consciousness Studies*, **5**(5 - 6), 592 – 610.
- Szabó Gendler, T. and Hawthorne, J., editors (2002). *Conceivability and Possibility*. Clarendon Press, Oxford.
- Thomson, J. J. (1997). People and their bodies. In J. Dancy, editor, *Reading Parfit*. Blackwell, Oxford.
- van Inwagen, P. (1990). Four-dimensional objects. *Nous*, **24**(2), 245 – 255.
- Velleman, D. (1996). Self to self. *The Philosophical Review*, **105**(1), 39 – 76.
- von Stechow, A. (1982). Structured propositions. Unpublished manuscript.
- von Stechow, A. (2002). Binding by verbs: Tense, person and mood under attitudes. Unpublished manuscript.
- von Stechow, A. (2003). Feature deletion under semantic binding: Tense, person, and mood under verbal quantifiers. In M. Kadawaki and S. Kawahara, editors, *Proceedings of NELS 33*.
- Walton, K. (1990). *Mimesis as Make-Believe*. Harvard University Press, Cambridge, MA and London.
- Williams, B. (1966). Imagination and the self. In Williams (1973), pages 26 – 45.
- Williams, B. (1970). The self and the future. In Williams (1973), pages 161 – 180.
- Williams, B. (1973). *Problems of the Self : Philosophical Papers 1956-1972*. Cambridge University Press, Cambridge.
- Zimmerman, D. (1998). Criteria of identity and the ‘identity mystics’. *Erkenntnis*, **48**, 281 – 301.