

C O U N T E R F A C T U A L S
A N D
(C O U N T E R -) I D E N T I T Y

T H E I D E N T I T Y C R I S I S O F
“ I F I W E R E Y O U ”

Carina Kauf

Prof. Dr. Hedde Zeijlstra
Prof. Dr. Markus Steinbach

**C O U N T E R F A C T U A L S
A N D
(C O U N T E R -) I D E N T I T Y
T H E I D E N T I T Y C R I S I S O F
“ I F I W E R E Y O U ”**

**Masterarbeit „Master of Education“
Lehramt an Gymnasien
an der Georg-August-Universität Göttingen**

vorgelegt am 30. Mai 2017

von Carina Kauf

aus Varel

Erstgutachten: Prof. Dr. Hedde Zeijlstra

Zweitgutachten: Prof. Dr. Markus Steinbach



**SOZIALWISSENSCHAFTLICHE FAKULTÄT DER
GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN**

Abstract

In this thesis, I focus on the semantic analysis of “counteridenticals” (Goodman 1984), a subclass of counterfactuals whose antecedents identify two inherently incompatible entities with each other, e.g. “If I were you, I would VP.” As a novel contribution, I argue that counteridenticals are grammatically ambiguous between an ‘advice’ reading (If I were you, I would buy the blue dress. I like it much better than the red one.) and an ‘imagine’ reading (I’m so jealous of you right now; If I were you, I would already be done with all of my papers and could enjoy the weather. Instead, I am stuck at my desk.). For this purpose, I adduce morphosyntactic evidence from Korean, Polish and LIBRAS (=Brazilian Sign Language) as well as semantic/pragmatic arguments in form of constraints on pronoun reference and choice of pronouns. I show that neither this ambiguity nor complex reference patterns such as “If I_i were you $_j$, I_k would be sitting where you $_j$ are and I_k would be looking at me $_i$.” are predicted by existing theories of counterfactual meaning (Iatridou 2000; Ippolito 2013) and that they cannot be captured by existing proposals of the meaning of counteridenticals (e.g. Lakoff 1996; Malamud 2006; Kocurek 2016), either. As a solution, I suggest to exploit the structure’s similarity to dream reports (e.g. regarding the ambiguity of pronoun reference) and argue for a semantic analysis which makes use of concept generators in their realization as centered worlds after Percus and Sauerland (2003), and Ninan (2008), therefore. To this end, a redefinition of counteridentical propositions as “attitudinal objects” (Moltmann 2003) is necessary.

Key words: *Counteridenticals, counterfactuals, counteridentity, identity statements, counterparts, dream reports, centered worlds*

Acknowledgments

The best part about writing this thesis is the great people I got to meet and work with along the way. First of all, I am grateful to Hedde Zeijlstra for the valuable comments and constructive suggestions he provided at each stage of writing this thesis, and especially for taking the time to discuss my ideas with me on such a regular basis, first over Skype when I was still abroad and then in person. Furthermore, I would like to thank Markus Steinbach, whom I could always come to with any kind of question – thesis-related or not –, and who makes being interested in topics that lie outside of the curriculum’s usual scope so uncomplicated. I would like to express my very great appreciation to Kate Davidson who was my mentor during my stay at Harvard and who had been involved in my thesis from the very beginning. I thank Kate for her valuable support in how to approach this vast topic as well as for her never ending enthusiasm. Furthermore, I would like to thank the Linguistics Department of Harvard University for an inspiring and unforgettable research fellowship during which the contents of this thesis were developed. I want to thank Gennaro Chierchia and Kai von Stechow for interesting conversations, which not only helped me order my ideas but also left me with much new food for thought, and Roberta Pires de Oliveira for being as excited about this topic as I am. I further offer my special thanks to the members of the Meaning & Modality Lab of Harvard University, who never tired of talking to me about counteridenticals, who took part in my study, and who let me present my ideas in the Lab. I would also like to thank Sarah Zobel for interesting discussions and comments, as well as the members of the Oberseminar Linguistics of the University of Göttingen for their insightful questions and remarks. Lastly, I would like to thank everyone who proofread my thesis. All errors are, of course, mine.

Contents

Acknowledgments	ii
1 Introduction	1
2 The Puzzle: Lexical Ambiguity of <i>If I Were X?</i>	3
2.1 Introducing the Phenomenon	3
2.2 Evidence for Real Ambiguity	5
2.2.1 Morphosyntactic Evidence	6
2.2.1.1 Korean	6
2.2.1.2 Polish	8
2.2.1.3 LIBRAS	9
2.2.2 Semantic/Pragmatic Evidence	11
2.2.2.1 Interpretation of Pronouns	11
2.2.2.2 Choice of Predicates	14
2.3 Formulation of Research Question	18
3 Theories of Counterfactual Conditionals	19
3.1 Definition and Terminology	19
3.1.1 Conditionality	19
3.1.2 Counterfactuality	20
3.2 Discussion of Existing Theories	22
3.2.1 Past-as-Unreal Approach – Iatridou (2000)	22
3.2.2 Past-as-Past Approach – Ippolito (2013)	25
3.2.3 Application to Counteridenticals	29
3.3 Concluding Remarks	37
4 The Semantics of Identity Statements	40
4.1 Definition and Terminology	40
4.2 Problematization and Differentiation of the Notion ‘Identity Statement’	41
4.2.1 Identity Statements in Cross-World Contexts	42
4.2.2 Identity Statements in Counteridenticals	43
4.2.2.1 Symmetric Identity Statements – Pronoun Resolution	44
4.2.2.2 Asymmetric Identity Statements – Antecedent Relation	45
4.3 Concluding Remarks	48
5 Theories of Counteridentical Conditionals	50
5.1 A Combinatorial Analysis?	50
5.2 Counterpart Theories of Counteridenticals	51
5.2.1 Preliminaries: Lewis (1973)	52

5.2.2	Lakoff (1996)	55
5.2.3	Malamud (2006)	59
5.2.4	Kocurek (2016)	63
6	An Analysis of Counteridenticals in Terms of Dream Reports	70
6.1	Parallels Between Counteridenticals and Dream Reports	70
6.2	Discussion of Existing Proposals of Dream Reports	73
6.2.1	Percus and Sauerland (2003)	74
6.2.2	Ninan (2008)	76
6.3	Direct Transfer to Counteridenticals?	77
6.4	Proposal	80
7	Conclusion and Outlook	84
	References	87

List of Figures

1	Iatridou's approach to counterfactuals	25
2	Branching futures model	26
3	Ippolito's approach to conditionals	27
4	Lakoff's model of <i>If I were you</i> -conditionals	56

*“You never really understand a person
until you consider things from his point
of view – until you climb inside of his
skin and walk around in it.”*

HARPER LEE,
To Kill a Mockingbird

1 Introduction

The semantic analysis of counterfactual conditional sentences has been one of the prevalent topics in linguistics research for some time now. Besides working towards a theory which is able to capture the correct meaning of the different kinds of counterfactuals, as in any semantic investigation, the semanticists are also trying to upkeep the ideal of ‘Frege’s paradise,’ in which a structure’s meaning is always derived compositionally, from the meanings of its parts and the way these parts are syntactically combined (cf. Frege 1879). The thesis at hand joins this line of research by investigating the semantics of a subclass of counterfactual conditional sentences which has received only little attention in the literature so far, but which should be considered in quest of a unified analysis of counterfactual meaning: *If I were X*-conditionals (also dubbed ‘counteridenticals’ after Goodmann (1984)). As a novel contribution, it is argued that the antecedent clause *If I were X* is semantically ambiguous between two readings – an insight which is neither predicted by nor easily compatible with the existing theories of counterfactuality. It is the aim of this thesis to investigate the meaning of the two different interpretations of *If I were X*-conditionals and to analyze how the according semantic structures arise from the meanings of their parts, therefore.

Based on this agenda, the following structure has been developed for this thesis: The next chapter introduces the phenomenon which lies at the heart of this work; it argues for the existence of two different interpretations of the antecedent clause *If I were X*, first on the basis of their intuitive distinction in meaning (\Rightarrow chapter 2.1), and then by means of adducing arguments from the fields of morphosyntax, semantics and pragmatics (\Rightarrow chapter 2.2). After the formal ambiguity of the antecedent clause has been established, the last section of this chapter serves to explicate the aim of this thesis via formulating a specific research question as well as the subquestions which have to be answered in quest of its solution (\Rightarrow chapter 2.3).

Chapters 3 and 4 lay down the relevant theoretical background for a combinatorial analysis of *counter-identicals*. The two chapters are structured in a similar manner: After having given a definition of the respective notion, i.e. ‘counterfactual conditional’/‘identity statement’ (\Rightarrow chapters 3.1; 4.1), chapter 3 continues to sketch the two prevalent theories of counterfactual meaning (Iatridou 2000; Ippolito 2013) and discuss whether they are successful in capturing the semantics of counteridenticals in light of the findings of chapter 2 (\Rightarrow chapter 3.2); chapter 4, on the other hand, goes on to introduce the range of meanings identity statements have been made out to possess in the literature (\Rightarrow chapters 4.1-4.2.1) and to reveal where in the analysis of counteridenticals these contribute to the structure’s meaning (\Rightarrow chapters 4.2.2-4.3).

Chapters 5 and 6 make up the main part of this thesis; they join the previously discussed subcomponents in the semantic analysis of counteridenticals. Section 5.1 hereby

serves to show that a proposal of counterfactual meaning which simply combines either of the theories of counterfactual with those of identity statements does not yield the desired result – it can be regarded as a brief reminder as well as a formalization of the criticism raised in chapter 3.2, therefore. Due to the insufficiency of such an account, it becomes evident that some additional factor has to be considered in the analysis of *If I were X*-conditionals. The factor which has been suggested to do the job in the literature is cross-world rigidity of entities. Based on Lewis' (1973) counterpart theory of cross-world rigidity, 'counterpart theoretical' proposals of *If I were X*-conditionals have been put forward in the frameworks of mental space theory (Lakoff 1996) and formal semantics (Malamud 2006; Kocurek 2016). Ultimately, this thesis proposes to also take another factor into consideration in the analysis of *If I were X*-conditionals, i.e. their connection to dream reports. In order to pave the way for such an analysis, chapter 5.2 goes at length through the different counterpart proposals to show that none of them are able to capture the entire set of relevant data.

Chapter 6 suggests to analyze *If I were X*-conditionals in a similar way to dream reports. In order to motivate such a theory, in a first step, several syntactic and semantic parallels will be drawn between the two independent areas of linguistic research (\Rightarrow chapter 6.1). Afterwards, two different proposals for the meaning of dream reports (Percus and Sauerland 2003; Ninan 2008) will be sketched in order to get acquainted with the relevant formal toolkit (\Rightarrow chapter 6.2). Chapter 6.3 discusses the technical problems which arise from a transfer of a theory of dream reports to counterfactuals, and suggests to remedy these by means of Moltmann's (2003) view of propositions as attitudinal objects. Chapter 6.4 contains the formal proposal of counterfactual meaning which this thesis puts forward. It generalizes Moltmann's theory and combines it with a centered worlds-analysis of counterfactuals in the fashion of dream reports, whereby the denotation of the center deviates between the two kinds of *If I were X*-conditionals.

Chapter 7 serves to conclude and to motivate future research with regard to the topic of this thesis by recapitulating some important issues that were raised and have left questions for further investigation in this work, as well as via introducing further associated questions, which have not found their way into this work.

2 The Puzzle: Lexical Ambiguity of *If I Were X*?

What is meant by the ‘ambiguity of the clause *If I were X*’? In how many ways can this phrase be interpreted? And what line of argumentation justifies the claim of a grammatical instead of a pragmatic distinction of these different readings? These are the questions that are set out to be answered in the following chapter.

2.1 Introducing the Phenomenon

In order to illustrate the proposed ambiguity, consider the following set of sentences.

- (1)
 - a. *If I were you*, I’d buy the blue dress. I like it much better than the red one.
 - b. *If I were Stephen Hawking*, I would’ve insisted on a speaking device with a British accent. It surprises me that he didn’t.
- (2)
 - a. I’m so jealous of you right now; *If I were you*, I would already be done with all of my papers and could enjoy the weather. Instead, I am stuck at my desk.
 - b. *If I were Stephen Hawking*, I would be the author of “A Brief History of Time.”

Upon reading these sentences, it intuitively becomes clear that there exists a marked difference in meaning between the expression *If I were X* as used in (1) and the one as used in (2). This difference in interpretation arises from the fact that in (1), the speaker of the utterance counterfactually entertains the idea of *being in somebody else’s position* – i.e. considering things from somebody else’s point of view but keeping one’s own mindset –, whereas in (2), the speaker of the utterance is counterfactually imagining what it would be like *were he¹ actually somebody else* – i.e. were he to literally ‘climb inside of someone else’s skin and walk around in it.’ Sentences like (1-a) are usually understood as advice and will be referred to as the ‘advice’ use of *If I were X* in this work, therefore.

The same label will be adopted for sentences similar to (1-b), even though this reference might not come about as intuitively: It is commonly argued that an advisability reading of counterfactuals can only arise when the consequent is interpreted as referring to the future, since it seems pragmatically odd to advise someone to do something in retrospect (cf. Declerck and Reed 2001: 272). To the best of my knowledge, it is further commonly assumed that advice can only be given in a speaker-addressee-context and, thus, should have the form “If I were you, . . .” Under these premises, it seems odd that sentences of the form “If I were X, . . .,” where X is not the addressee of the conversation, should be interpreted as advice. Nevertheless, contrary to these assumptions, I argue that it makes sense to interpret them along the same lines and, thus, also refer to them in the same way for the purpose of this paper. This hypothesis is based on the fact that there

¹ Note that since most of the examples in this thesis include female characters, the speaker and the addressee will be denoted by male pronominals in order to avoid confusion in the analyses.

clearly exists a point in time prior to the utterance time at which the speaker could have uttered the sentence “I would insist on a speaking device with a British accent.” as an advice to Stephen Hawking under the notion’s usual interpretation. As a consequence, I suggest that sentences like (1-b) can be paraphrased as “If Stephen Hawking had been here before he accepted the device, I would have said to him: ‘If I were you, I would insist on a speaking device with a British accent,’” which satisfies the restrictions for an ‘advice’ reading under the usual assumptions.

In contrast to (1-a) and (1-b), sentences similar to those subsumed under (2) can be argued to demand a higher level of imagination, because in employing them, one does not only have to change one’s point of view but to actually abandon one’s own body, as well. Based on this argumentation, the usage of the antecedent clause in (2) will be referred to as its ‘imagine’ reading. Lastly, conditional sentences whose antecedents are of the form *If I were X* will be dubbed either ‘*If I were X*-conditionals’ or ‘counteridentical counterfactuals’/‘counteridenticals,’ following Goodman (1984), in this work. The notions of conditionality/counterfactuality and counteridentity will be explained in detail in chapters 3 and 4, respectively.

The fact that the two applications of one and the same antecedent clause introduced above are indeed distinct becomes evident when applying a simple empirical test: We try to follow each of the prefatory sample sentences with the phrase “That’s/That would’ve been my advice for X,” which explicitly requires the preceding sentence to have an ‘advice’ interpretation, or change the sentence’s condition to “In X’s shoes” instead of “If I were X,” which – as a structure known to be restricted to giving advice in English – also tests for this reading. In doing so, we see that while such a continuation/substitution works fine for the sentences listed under (1), it renders an infelicitous result for (2-a) and (2-b).

- (3) a. If I were you, I’d buy the blue dress. [*I like it much better than the red one.*]
 - (i) If I were you, I’d buy the blue dress. That’s my advice for you.
 - (ii) In your shoes, I’d buy the blue dress.
- b. If I were Stephen Hawking, I would’ve insisted on a speaking device with a British accent. [*It surprises me that he didn’t.*]
 - (i) If I were Stephen Hawking, I would’ve insisted on a speaking device with a British accent. That would’ve been my advice for him.
 - (ii) In Stephen Hawking’s shoes, I would’ve insisted on a speaking device with a British accent.
- (4) a. [*I am so jealous of you right now;*] If I were you, I would already be done with all of my papers and could enjoy the weather. [*Instead, I am stuck at my desk.*]
 - (i) *If I were you, I would already be done with all of my papers and could

- enjoy the weather. That's/That would've been my advice for you.
- (ii) *In your shoes, I would already be done with all of my papers and could enjoy the weather.
- b. If I were Stephen Hawking, I would be the author of "A Brief History of Time."
 - (i) *If I were Stephen Hawking, I would be the author of "A Brief History of Time." That's/That would've been my advice for him.
 - (ii) *In Stephen Hawking's shoes, I would be the author of "A Brief History of Time."

Further evidence for this contrast in interpretation is provided by the subsequent utterance, which can be felicitously followed by either of the two possible continuations, a. and b.

- (5) A: If I were you, I would buy the short, blue dress.
 - a. B: No, you wouldn't. You hate the color blue. You're just saying this because you know that I like it best.
 - b. B: No, you wouldn't. Then you'd know how insecure I am about my legs, so you'd always buy a long one to cover them up.

What is crucial about this example is that (5-a) questions the sincerity of A's advice-giving, i.e. whether A would really act the way as proclaimed in B's situation, whereas (5-b) serves as an explanation as to why B does not take A's advice – it asserts new information about B, which would have to be considered by A in order to give a 'good' piece of advice. Hence, the example illustrates that B interprets A's utterance *If I were you* differently in the two contexts: figuratively (= counterfactually assuming to be in A's situation) in (5-a) and literally (= counterfactually assuming to be, in fact, A) in (5-b).

2.2 Evidence for Real Ambiguity

The question which immediately presents itself from a linguistics point of view upon noticing such an ambiguity is the following: Is the construction *If I were X* ambiguous between an 'advice' and an 'imagine' reading or is this difference brought about by the context? More formally speaking, this translates into: Is the difference in interpretation a grammatical or a pragmatic issue? As has been stated in the beginning of this thesis, this work argues in favor of the former underlying reason, i.e. that there exists a *grammatical* difference between the two applications of the structure. In the following, two types of arguments will be adduced in order to support this hypothesis: Cross-linguistic, morphosyntactic evidence, and arguments from the field of semantics/pragmatics.

2.2.1 Morphosyntactic Evidence

The following morphosyntactic arguments have emerged as the result of a small cross-linguistic pilot study ($n = 8$) concerning the use and interpretation of the phrase *If I were X* that I conducted during my stay at the Harvard University Department of Linguistics. In this rather informal survey, the subjects were given a list of ten *If I were X*-conditionals as well as contexts of usage for each one of them, and were asked to translate the sentences into their mother tongue. The list included sentences of both proposed usages of the antecedent as well as combinations thereof (e.g. (11)). Whereas several languages, e.g. German, French and Romanian, have been found to coincide with English in that they do not mark the difference in use morphologically, a distinct behavior has been observed for Korean, Polish and LIBRAS (= Língua Brasileira de Sinais, Brazilian Sign Language), which will be analyzed in turn in the following.

2.2.1.1 Korean

In Korean, counterfactuality arises via a conversational implicature (cf. Han 1996: 2). According to Han, this implicature is drawn when a conditional sentence uses past-tense morphology in its antecedent and future-tense morphology in its consequent (cf. *ibid.*: 5). Extending Han's morphological discussion, the following four grammatical structures have been approved for Korean present counterfactuals by Ahn (p.c.). In view of the initial question of this survey, these structures were then judged with respect to their validity for the two uses of *If I were X*. The subsequent schema provides a summary of the study's findings:

- | | |
|-------------------------------------|--------------------------|
| 1. [...V-Past...if] [...V-Fut] | ['Advice' ✓/'Imagine' ✓] |
| 2. [...V-Past...if] [...V-Past-Fut] | ['Advice' ✓/'Imagine' ✓] |
| 3. [...V-Past...if] [...V-Past] | ['Advice' ✓/'Imagine' ✓] |
| 4. [...V-Past...if] [...V-Pres] | ['Advice' ✓/'Imagine' ✗] |

(Han 1996, extended by Ahn (p.c.)²)

What is notable about this data for the purpose of this paper is that the use of the PRES-IND in the consequent forces an 'advice' reading, while all other structures can be used ambiguously between the two suggested interpretations. This restriction in the distribution of the morphological structure 4. provides a first, subtle piece of evidence for the case of lexical ambiguity of the term *If I were X*. To illustrate this distinction more clearly, consider the following set of example sentences, provided by Ahn (p.c.), where (6) presupposes a context in which the speaker proclaims what he would do if he were in

²For reasons of clarity and comprehensibility, this analysis has been limited to Korean present counterfactuals. For an in-depth analysis of Korean past counterfactuals, the reader is encouraged to consult Han (1996).

Mary's situation – he would go into the sea, even though she might not –, and (7) could, for example, be uttered in a scenario in which the speaker has just received a message with a picture showing Mary going into the sea at this moment and is now fantasizing about being her.

- (6) If I were Mary, I would go in the sea right now. [‘Advice’]
- a. Nay-ka Mary-i-ess-tamyen, cikum-ccum pata-ey teleka-l.kess-ita.
I-NOM Mary-be-PAST-if around.now ocean-DAT enter-FUT-IND
‘If I were Mary, I would go in the sea right now.’
- b. Nay-ka Mary-i-ess-tamyen, cikum-ccum pata-ey teleka-ss-ul.kess.ita.
I-NOM Mary-be-PAST-if around.now ocean-DAT enter-PAST-FUT-IND
‘If I were Mary, I would go in the sea right now.’
- c. Nay-ka Mary-i-ess-tamyen, cikum-ccum pata-ey teleka-ss-ta.
I-NOM Mary-be-PAST-if around.now ocean-DAT enter-PAST-IND
‘If I were Mary, I would go in the sea right now.’
- d. Nay-ka Mary-i-ess-tamyen, cikum-ccum pata-ey teleka-n-ta.
I-NOM Mary-be-PAST-if around.now ocean-DAT enter-PRES-IND
‘If I were Mary, I would go in the sea right now.’
- (7) If I were Mary, I would go in the sea right now. [‘Imagine’]
- a. Nay-ka Mary-i-ess-tamyen, cikum-ccum pata-ey teleka-l.kess-ita.
I-NOM Mary-be-PAST-if around.now ocean-DAT enter-FUT-IND
‘If I were Mary, I would go in the sea right now.’
- b. Nay-ka Mary-i-ess-tamyen, cikum-ccum pata-ey teleka-ss-ul.kess.ita.
I-NOM Mary-be-PAST-if around.now ocean-DAT enter-PAST-FUT-IND
‘If I were Mary, I would go in the sea right now.’
- c. Nay-ka Mary-i-ess-tamyen, cikum-ccum pata-ey teleka-ss-ta.
I-NOM Mary-be-PAST-if around.now ocean-DAT enter-PAST-IND
‘If I were Mary, I would go in the sea right now.’
- d. *Nay-ka Mary-i-ess-tamyen, cikum-ccum pata-ey teleka-n-ta.
I-NOM Mary-be-PAST-if around.now ocean-DAT enter-PRES-IND
Intended: ‘If I were Mary, I would go in the sea right now.’

Besides this morphological distinction, yet another difference can be observed between the ‘advice’ and ‘imagine’ use of *If I were X* in Korean: the evidential marker *-ney* can only be used felicitously when aiming at the latter interpretation (p.c. Ahn).

- (8) a. Nay-ka Mary-i-ess-tamyen, cikum-ccum pata-ey teleka-kess-ney.
I-NOM Mary-be-PAST-if around.now ocean-DAT enter-NOM-EVI
‘[I infer that] If I were Mary, I would go in the sea right now.’
- b. Nay-ka Mary-i-ess-tamyen, cikum-ccum pata-ey teleka-ss-kess-ney.
I-NOM Mary-be-PAST-if around.now ocean-DAT enter-PAST-NOM-EVI
‘[I infer that] If I were Mary, I would go in the sea right now.’

Even though this further points towards an underlying grammatical instead of a pragmatic

issue, due to the constraints of this thesis, this observation will be left without further explanation; a more formal analysis of the distribution of *-ney* in this context is instead suggested as topic for future research.

Whereas the grammatical distinction between the two interpretations of *If I were X* in Korean is rather subtle, in Polish and LIBRAS it becomes more explicit, since both languages make use of different constructions for the counterfactual's antecedent in order to distinguish between the two readings.

2.2.1.2 Polish

In Polish, for example, one of the following two phrases is used when introducing a conditional sentence of the form “If I were you, ...”:

1. Na Twoim miejscu, ...
On your spot, ...
'In your spot, ...'
 2. Gdybym był tobą, ...
I be-PAST you, ...
'If I were you, ...'
- (p.c. Fuchs)

Crucially, it is not accepted to give advice using the antecedent listed under 2.; in this case, the construction “Na Twoim miejscu, ...” has to be used. The phrase “Gdybym był tobą ...,” on the other hand, is employed when aiming at an interpretation in which the speaker is actually imagining to be the addressee, i.e. in phrases similar to the prefatory sample sentences in (2) (*ibid.*).³ In order to illustrate the two antecedents' behavior in ‘advice’ and ‘imagine’ contexts, consider examples (9) and (10).

- (9) CONTEXT: Talking to someone who actually is from the US
If I were you, I'd be from the US.
- a. Gdybym był Tobą, byłabym Amerykanką.
If-1sg be-PAST you.instr, would.be-1sg American.instr.
'If I were you, I'd be from the US.'
 - b. *Na Twoim miejscu, byłabym Amerykanką.
On your spot, would.be-1sg American.instr.
Intended: 'If I were you, I'd be from the US.'

- (10) If I were you, I'd buy the blue dress.

³ As with any language, it could, of course, be the case that the copula ‘być’ (= to be) has a different semantics than its English correlate. If the copula's distribution were restricted to true identity, for example, an ‘imagine’ reading would arise as a natural consequence from its usage and the distinction of the two antecedent meanings would not seem surprising. From studies of the distribution of Polish copulas, like Citko (2008) or Bondaruk (2013), however, we know that *być* is not only used in equatives (cf. section 4), which speaks in favor of such a distinction.

- a. Na Twoim miejscu, kupiłaby-m niebieską sukienkę.
On your spot, buy.cond-1sg blue.f.acc dress.f.acc.
'If I were you, I'd buy the blue dress.'
- b. *Gdybym był Tobą, kupiłabym niebieską sukienkę.
If-1sg be-PAST you.instr, buy.cond-1sg blue.f.acc dress.f.acc.
Intended: 'If I were you, I'd buy the blue dress.'

(p.c. Fuchs)

An interesting implication which arises from this distinction is that in Polish, consequents of the two different classes of interpretations cannot be combined as they each require their own introductory phrase (*ibid.*). This stands in contrast to languages like English, where such a combination is possible.⁴

- (11) If I were you(=Michelle Obama), I'd live in the White House (because she does) and I(=the speaker)'d love it (e.g. because I have always wanted to live in such an impressive building). ✓

2.2.1.3 LIBRAS

As previously mentioned, Brazilian Sign Language distinguishes between the 'advice' and the 'imagine' reading of *If I were X* in a similar manner to Polish. In LIBRAS, too, there exist two distinct structures for the different uses of the counterfactual antecedent, which, here, are defined by the presence or absence of the morpheme INCORPORATE. The sign INCORPORATE, where both hands are held in a vertical line in a manner such that only the thumb and the index finger of each hand touch, can be roughly translated as 'climbing into the skin of X'. It consists of the described handshape as well as a variable component, i.e. its indexical movement, which signals who is taking over whose role by figuratively 'taking' someone's role – the starting point of the sign – and 'putting it onto' somebody else – the ending point of the sign. If the sign INCORPORATE is present, an 'imagine' reading emerges; otherwise, the counterfactual is interpreted as a counsel.

- (12) If I were Michelle Obama, I'd be from the US.
- a. $\overline{\text{FS(SI) IX(SELF) FS(MICHELLE OBAMA) INCORPORATE, IX(SELF) EUA}}$
If 1-sg Michelle Obama incorporate, 1-sg USA
'If I were Michelle Obama, I'd be from the US.'
- b. $\overline{* \text{FS(SI) IX(SELF) FS(MICHELLE OBAMA), IX(SELF) EUA}}$
If 1-sg Michelle Obama, 1-sg USA

⁴Likely, the distinction between minimal pairs like (11) and the sentence "If I were you (=Michelle Obama), I'd live in the White House (because she does) and I'd love it (because she does)." is realized by prosodic means. Due to the constraints of this thesis, however, an investigation of this claim is left to future research.

Intended: ‘If I were Michelle Obama, I’d be from the US.’,

where FS = fingerspell, and IX = index.

(13) If I were you, I’d buy the blue dress.

a. $\frac{\text{eyebrows-raised}}{\text{FS(SI) IX(SELF) IX(YOU)}, \text{IX(SELF) BUY BLUE DRESS.}}^{\text{head-nod}}$
 If 1-sg 2-sg, 1-sg buy blue dress.
 ‘If I were you, I’d buy the blue dress.’

b. $\frac{\text{eyebrows-raised}}{* \text{FS(SI) IX(SELF) IX(YOU) INCORPORATE}, \text{IX(SELF) BUY BLUE DRESS.}}^{\text{head-nod}}$
 If 1-sg 2-sg incorporate, 1-sg buy blue dress.
 Intended: ‘If I were you, I’d buy the blue dress.’

(p.c. Quadros)

Even though Polish and LIBRAS share the property that they both have designated structures for each of the two suggested uses of *If I were X*, given their difference in modality, certain dissimilarities arise: It has been argued that in Polish the different kinds of consequents cannot be combined as they both have their own fixed introductions which cannot be joined. In LIBRAS, by contrast, such a ‘role shift’ can easily be implemented in the middle of a sentence via the morpheme $\overline{\text{BUT}}^{\text{shift-back}}$. What this sign does is cancel the transfer of roles which has previously been introduced by means of the $\overline{\text{INCORPORATE}}$ morpheme. Interestingly, the shift is not achieved via ‘reversing’ the previous morpheme – this would yield a reading in which the addressee is taking over the speaker’s role – but rather in form of the commonly known strategy *role shift* (=rs), in which “the signer breaks eye gaze and/or moves his body to signal that the words used belong to somebody else” (Davidson 2014: 2). Thus, instead of glossing $\overline{\text{BUT}}^{\text{shift-back}}$, which has been chosen for reasons of clarity, the more common notation would have been $\overline{\text{BUT}}^{\text{rs}}$.⁵ Given the ability of this morpheme to transfer individuals back into their proper roles, sentences like (14) are possible in LIBRAS – as they are in English but not in Polish (see example (11)).

⁵The possibility to cancel the morpheme $\overline{\text{INCORPORATE}}$ by means of $\overline{\text{BUT}}^{\text{rs}}$ has led to a further interesting insight with regard to a different field of research, i.e. the analysis of demonstrations in sign language (cf. e.g. Davidson 2014). Davidson suggests that in sign language, role shift is “the closest equivalent to English ‘be like’”, and, as such, is responsible for the introduction of demonstrations (24). What has been discovered in this study is that there does not exist stacking of the morpheme $\overline{\text{INCORPORATE}}$ and the strategy role shift in LIBRAS. This means that in the minimal pair “If I were you, I’d be all like ‘[in a squeaky voice] Ahhh!’ now (because I am scared of spiders and you are looking at one).” and “If I were you, I’d be all like ‘[in a squeaky voice] Ahhh!’ now (because that’s what you would do in this situation).”, where the $\overline{\text{INCORPORATE}}$ morpheme would only be used in the latter example, the strategy role shift would only be used in the former one. This seems surprising given that the sign/the strategy are not interchangeable and that the former has not been established as an introducer of demonstrations. An intuitive explanation for this phenomenon, provided by Quadros (p.c.), is based on the fact that in the former case “the speaker is already in the addressee’s role and no extra shift is necessary to demonstrate his behavior, therefore.” Again, this topic is left to others to investigate.

- (14) $\overline{\text{FS(SI) IX(SELF) FS(MICHELLE OBAMA) INCORPORATE,}}$ ^{eyebrows-raised}
 If 1-sg Michelle Obama incorporate,
 $\overline{\text{IX(SELF) LIVE WHITE-HOUSE BUT IX(SELF) LOVE.}}$ ^{head-nod shift-back head-nod}
 1-sg live White House but 1-sg love.
 ‘If I were Michelle Obama, I’d live in the White House (because she does) and I(=the speaker)’d love it.’
 (p.c. Quadros)

All things considered, this section has shown that the data from Korean, Polish and LIBRAS provides strong morphosyntactic arguments in favor of a grammatical and against a pragmatic distinction that underlies the different interpretations of the construction *If I were X*. In order to further strengthen this argument, the next section will discuss additional evidence from the field of semantics.

2.2.2 Semantic/Pragmatic Evidence

The main claim of this subchapter is that there are two ways to semantically and pragmatically distinguish between the two proposed meanings of the clause *If I were X*. First, it will be shown that the first person pronouns in the consequent of a counterfactual cannot be interpreted arbitrarily but that there exist patterns of interpretation which correlate with the ‘advice’ and the ‘imagine’ reading of the antecedent. Afterwards, it will be argued that the usage of a particular class of predicates in the consequent clause further serves to disambiguate between the two readings, and that the speaker’s as well as the addressee’s properties impose restrictions on the choice of verbs which can be used felicitously for the different usages of *If I were X* in any case.

2.2.2.1 Interpretation of Pronouns

The first thing that is worth noting in this context is that the role shift in (14) unmistakably proves that the two *I*’s in the counterfactual’s consequent refer to two different individuals – possibly the subject and the object of the antecedent. Had the speaker wanted to refer to one and the same first person in both cases, the employment of the role-shift strategy would have been superfluous, or rather, misleading. If we investigate this issue further, it can be shown that the interpretation of the first person singular pronouns in the consequent of a counterfactual sentence constitutes a semantic piece of evidence in favor of the lexical ambiguity of the antecedent clause *If I were X*. In the following, this claim will be explained more thoroughly.

Before we proceed with the discussion, however, let us introduce some terminology which facilitates the analysis. From the subsequent data, we will see that the use of the

first person singular pronoun in the consequent of a counterfactual can have three different interpretations, which I call ‘ $I_{as S}$,’ ‘ $I_{as X}$ ’/‘ $I_{as A}$,’ and ‘ I_S .’ The first term, ‘ $I_{as S}$,’ will be applied to refer to the subject of an ‘advice’ interpretation of *If I were X*. Hence, the term is used when the speaker is describing an event in which he is counterfactually imagining to be in X’s situation but is keeping his own internal identity. In addition to $I_{as S}$, we also need a way of referring to the subject of an ‘imagine’ reading of the antecedent clause. To this end, the initials of the person that the speaker is imagining to be ‘climbing inside of the skin’ of will be used as a subscript; this notation serves to illustrate that the speaker is making a claim about himself as that person. When the speaker is imagining to be the addressee, the term ‘ $I_{as A}$ ’ will be used, accordingly. Lastly, we will see that there are also situations in which the speaker wants to refer back to his actual self, i.e. his self in the real world and neither merely his personality within someone else’s body nor himself as an entirely different person; this interpretation of the pronoun will be captured by the term ‘ I_S .’ This last relation of reference is usually indicated by co-indexation; in order to preserve consistency of nomenclature, however, the introduced subscript notation will be used in this work.⁶

With this notation in place, let us consider the following sentence:

(15) If I were you, I’d feed my cat every day.

It can be argued that the most salient interpretation of this sentence is one where the speaker advises the addressee to feed his, i.e. the addressee’s own, cat every day. A possible scenario in which this sentence with this intended meaning could be uttered is one in which the addressee has just gotten a new cat and has asked the speaker for advice regarding its care. Such a parsing can be schematized in the following way:

(16) If I were you, I’d feed my cat every day.

$I_{as S}$ $I_{as A}$

Here, the speaker first employs the – by now well-known – strategy to give advice, i.e. projecting his personality into the addressee’s situation and saying what he would do were he in that scenario. This yields an $I_{as S}$ -interpretation for the first first person pronoun in the consequent. The ‘my,’ however, is to be interpreted differently: In the real world,

⁶It has been brought to my attention by Zobel (p.c.) that it could also be the case that these pronouns do not refer to different individuals but rather to the same person in different roles. Such an analysis is consistent with the subscript notation of ‘as X’ used in this work, which hints at one’s overtaking of a role rather than there being two distinct individuals. Even though I am much looking forward to Zobel’s (forthcoming) analysis of *as*-phrases that could provide the basis of an analysis of counterfactuals along these lines, in this thesis, I will follow the line of literature on cross-world rigidity to this date in assuming that the combination of one person’s internal identity and someone else’s body, or even a person’s overtaking of an entirely different personality, cannot be assumed to still yield the original person as a referent in the actual world.

the cat which is the topic of the clause indeed belongs to the addressee; therefore, by employing the use of the first person in this case, an ‘imagine’ reading arises, which leads to its being interpreted as an $I_{as A}$. One could argue that instead of using the denotation $I_{as A}$ for explicating the reference of ‘my,’ one could also have used $I_{as S}$, since that I which possesses the speaker’s internal and the addressee’s external properties will also be the owner of the addressee’s cat. Nevertheless, I argue that it is always the strongest identification counterpart possible which determines the denotation of a pronoun, which – in cases where a property already holds of the addressee in the actual world – yields the $I_{as A}$ -denotation.

Even though the usage of (15) discussed above might be the most common use of this sentence, there exist three more possible interpretations of it. Consider, for example, the following situation: I am about to go on a holiday and I have asked you to feed my cat while I am gone. You have promised to do so, but I know that you tend to forget things, which – of course – would make me furious. Before I leave I, therefore, tell you:

(17) If I were you, I’d feed my cat every day.

$I_{as S}$ I_S

(Otherwise I’m gonna get really upset with you!)

In contrast to example (16), in (17) the speaker is no longer talking about the addressee’s cat that is to be fed, but rather about his own. As a result, the ‘my’ can be interpreted as an actual reference of the speaker, without any counterfactual imagining taking place, which leads to an I_S -interpretation. What (16) and (17) have in common is that both of them could be followed by “That would be my advice for you,” which implies an ‘advice’ interpretation of the sentences’ antecedents. Therefore, an $I_{as S}$ interpretation emerges for the first usage of the first person pronoun in the consequent of this example.

In accordance with the previous data, there also exist analyses of this sentence which cannot be subsumed under the umbrella term ‘advice.’ These interpretations arise when the sentence is uttered in a situation in which the speaker imagines what he would do were he actually the addressee (and not himself) – feeding his own, i.e. after the transformation, the addressee’s, cat every day (18-a), or feeding his own, i.e. in the real world, the speaker’s, cat every day (18-b). It is left to the reader to think of possible scenarios for the respective interpretations (18-a) and (18-b).

(18) If I were you, I’d feed my cat every day.

a.	$I_{as A}$	$I_{as A}$
b.	$I_{as A}$	I_S

As an interesting aside, consider that it is not possible for the speaker to refer back to the actual addressee by any means of a first person pronoun application. In this case,

the usual reference ‘you’ has to be used. To illustrate this finding, let us have a look at example (19), which includes not only an ‘imagine’ interpretation of *I* but also references to both of the actual people introduced by the antecedent and, thereby, presents one of the most challenging examples for the structure’s semantical analysis:

- (19) a. If I were you, I’d be sitting where you are and I’d be looking at me.

$$\text{I}_{as A} \qquad \text{you}_A \qquad \text{I}_{as A} \qquad \text{I}_S$$
- b. *If I were you, I’d be sitting where I am and I’d be looking at me.

$$\text{I}_{as A} \qquad \text{I}_A \qquad \text{I}_{as A} \qquad \text{I}_S$$

To summarize, this analysis shows that the distribution of the variables $\text{I}_{as S}$, $\text{I}_{as A}$, and I_S is not random with respect to the two possible readings of the antecedent clause *If I were X*: When an ‘advice’ reading is aimed for, the first *I* in the consequent clause is always interpreted as $\text{I}_{as S}$, while when an ‘imagine’ reading is intended, the subject-*I* can never be interpreted in this way. In such a case, the pronoun is usually understood as $\text{I}_{as A}$ (cf. (18)). Nevertheless, the following example shows that I_S can also be the subject of the consequent under the ‘imagine’ interpretation of *If I were X*:

- (20) CONTEXT: I am looking at you.
- a. If I were you, I’d be looking at me/myself right now.

$$\text{I}_S \qquad \text{I}_{as A}$$

Even though the systematic distribution of the $\text{I}_{as S}$ - vs. non- $\text{I}_{as S}$ interpretation of the first usage of the first person pronoun in a counterfactual distinctly marks an ‘advice’ versus an ‘imagine’-reading of *If I were X*, there can still exist ambiguity of these sentences. These equivocal readings are the result of further usages of the first person pronoun in the conditional’s consequent, which can often be interpreted in different ways (cf. (16)/(17), (18-a)/(18-b)); in fact, if we investigate this issue further, it becomes evident that the more first person pronouns are used in the consequent of a counterfactual, the more interpretations may arise. In such cases of ambiguity, the context provides clues which help disambiguate between the different readings and arrive at the one intended by the speaker.

2.2.2.2 Choice of Predicates

Nonetheless, the context is not the only factor by means of which equivocal counterfactuals can be disambiguated: the choice of the predicates used in the conditional’s consequent can also fulfill this role. An interesting insight in this regard is constituted by the fact that, in contrast to sentence (15), no fourfold interpretation is available for sentence (21).

(21) If I were Michelle Obama, I'd live in the White House and I'd love it.

It has been attested that the only two acceptable readings of this sentence are those in which the first pronoun in the conditional's consequent is interpreted as not referring to the speaker in Michelle Obama's shoes but to him in the literal sense of *being* her. Hence, the following schema of interpretation arises:

If I were Michelle Obama, I'd live in the White House and I'd love it.

(a)	I_{as} M.O.	I_S	✓
(b)	I_{as} M.O.	I_{as} M.O.	✓
(c)	I_{as} S	I_S	✗
(d)	I_{as} S	I_{as} M.O.	✗

What this pattern proves is that there appear to be restrictions with regard to which things can be advised and which properties can be taken over from another person in order to arrive at an 'imagine' reading. In the following, we will see that the use of stative verbs in the consequent clause, for example, inevitably leads to an 'imagine' reading of the consequent, while the use of eventive predicates can lead to either of the two possible readings of *If I were X*. In these cases, however, pragmatic constraints apply which help disambiguate between the two readings and which I do not want to leave unmentioned, therefore, either. An investigation on the different classes of constraints will be the concern of the next paragraphs.

First, consider the 'advice' use of *If I were X*. Via choosing properties from different grammatical categories that are pragmatically associated either with the antecedent's subject and or with its object, and testing the emerging sentences' validity for the desired interpretation of *If I were X* in a trial-and-error manner, the scope of the advisable properties is sought to be established.⁷

(22) 'Advice' use of *If I were X*

- a. If I were Michelle Obama, I_{as} S'd tell my husband not to endorse Donald Trump.
 - b. If I were Michelle Obama, I_{as} S wouldn't be sad about moving out of the White House.
 - c. If I were Michelle Obama, I_{as} S'd try wake boarding. I just did and it was so much; she seems like she could use some cheering up.
-
- d. *If I were Michelle Obama, I_{as} S'd love eating nutella.
 - e. *If I were Michelle Obama, I_{as} S'd be able to play volleyball well.
 - f. *If I were Michelle Obama, I_{as} S'd be from Germany.

⁷I do not raise a claim to completeness with regard to this analysis.

- g. If I were Michelle Obama, I_{as} S 'd move to Spain.
- h. *If I were Michelle Obama, I_{as} S 'd live in Spain.

- i. *If I were Michelle Obama, I_{as} S 'd have two daughters.
- j. *If I were Michelle Obama, I_{as} S 'd be in love with the president.

Examine the first set of sentences, (24)-(22-f). The common denominator of these examples is provided by the fact that all of the features which the speaker attempts to advise are derived from his own set of defining properties. It is the speaker that does not like Trump and does not want him endorsed, therefore; it is the speaker that wouldn't be sad about moving out of the White House; it is the speaker that loves eating nutella, etc. Nevertheless, only the first three examples are accepted as 'advice' uses of the structure *If I were X*. What can be concluded from this juxtaposition of sentences is that one can only advise features which are not stative properties of oneself; You can counsel someone with regard to what you thinks he should do ((24), (22-c)) or feel (22-b), but you cannot give advice with respect to stative attitudes (22-d), abilities (22-e) and innate properties (22-f).

The fact that only agentive verbs can be used to give advice, except for when talking about attitudes, becomes even more evident when consulting the examples (22-g) and (22-h) (cf. Declerck and Reed 2001: 272). Whereas the former is interpreted as a consult, the use of a stative verb in the latter's consequent renders an 'imagine' reading. Lastly, the examples (22-i) and (22-j) prove that the 'advice' use of *If I were X* only arises when the speaker is suggesting properties that are associated with himself. If properties of the antecedent's object are the topic of the consequent, an 'imagine' reading emerges.

Let us turn our attention to the 'imagine' reading of *If I were X* and employ the same strategy. Thereby, we might arrive at the following set of test sentences:

- (23) 'Imagine' use of *If I were X*
- a. If I were Michelle Obama, I_{as} M.O. 'd live in the White House.
 - b. If I were Michelle Obama, I_{as} M.O. wouldn't be working on my thesis right now.
 - c. If I were Michelle Obama, I_{as} M.O. 'd be in love with the president.

 - d. *If I were Michelle Obama, I_{as} M.O. 'd live on Oxford Street.
 - e. *If I were Michelle Obama, I_{as} M.O. 'd be working on my thesis right now.
 - f. *If I were Michelle Obama, I_{as} M.O. 'd love eating nutella.

 - g. *If I were Michelle Obama, I_{as} M.O. 'd be able to ride a tiger.

The consequents of (23-a)-(23-c) include properties which hold of the object of the

antecedent, here: Michelle Obama. Regardless of whether eventive or stative verbs are used, as long as the described property is believed to be true of Michelle Obama by the speaker, an ‘imagine’ reading is available for these counterfactuals. This stands in marked contrast to examples (23-d)-(23-f), in which the consequent describes features which hold of the subject of the counterfactual, i.e. the speaker of the utterance. Furthermore, example (23-g) shows that a property which neither the antecedent’s subject nor its object possesses cannot be used for an ‘imagine’ use of *If I were X*, either. In conclusion, the picture which emerges is that in order to arrive at an ‘imagine’ interpretation of the structure, the properties which provide the topic of the conditional’s consequent have to hold of exactly one person introduced by the antecedent, namely of its object.

In order to show that not only stative attitudes can be taken over by the speaker in the case of the ‘imagine’ use of *If I were X*, consider the following situation, pointed out to me by von Stechow (p.c.): Kai von Stechow is reading a syntax paper. He does not really enjoy it, but he knows that this is exactly the kind of literature his colleague, the syntax professor David Pesetsky, would like. In such a situation, Kai could felicitously utter the following sentence, imagining what it would be like to read the paper were he not himself but rather his colleague.

(24) If I were Pesetsky, I_{as D.P.} would enjoy this much more!

All in all, the following validity restrictions emerge for the two uses of *If I were X*. An ‘advice’ use is available whenever the speaker advises the addressee something that he thinks the listener should do or feel. The properties which can be advised need to be properties which a person can take influence on, a restriction which rules out the advisability of stative properties. Furthermore, the advised properties must either hold of the speaker in the actual world or hypothetically be true of the speaker were he put into the object’s position.⁸Note that these restrictions do not include the fact that the property the speaker advises has to be *false* of the listener. In , for example, the speaker can advise Michelle Obama to tell her husband not to endorse Trump, even though she might already do that independently of the speaker’s advice.

The ‘imagine’ use of *If I were X* seems to be somewhat less restricted. Such a reading arises if the property described in the consequent of the conditional sentence is true for the object of its antecedent. In this case, no restriction emerges with regard to the property’s stativeness/eventivity, since no influence needs to be taken on the property the speaker counterfactually imagines to hold of himself; as long as the feature is associated with the

⁸In this work, cases in which the speaker of a sentence lies are being neglected. An example of such a falsehood has been given in (5), restated here as (i)

- (i) A: If I were you, I would buy the blue dress.
 B: No, you wouldn’t. You’d never buy a dress that expensive.

addressee, the structure can be used by the speaker to express what it must be like to ‘climb inside of the addressee’s skin’ and, thus, have one of the addressee’s designated properties be true of himself as well. As a consequence, this means that whenever a stative predicate appears in the consequent of a counterfactual conditional, an ‘imagine’ reading emerges.

2.3 Formulation of Research Question

In conclusion, all of the observations stated and discussed in this chapter prove that there exists a grammatical difference between the ‘advice’ and the ‘imagine’ use of the antecedent clause *If I were X* – a fact which, to the best of my knowledge, has been unacknowledged in the literature so far. This, in turn, means that semantic theories of counterfactual conditional sentences need to be able to a) account for both of the antecedent’s readings, and b) explain why, in some cases, only one of them can be applied felicitously. Hence, the aim of this thesis may be explicated in the following way:

Research Question

How do we achieve the different denotations of the pronouns in the ‘advice’ and the ‘imagine’ use of counterfactual conditionals, and, thereby, the different readings?

Associated sub-questions, which can be deduced from the preceding discussion, include:

Associated Sub-Questions

1. What is the relationship between the subject and the object of the conditional’s antecedent?
 - How is the incorporated identity statement (*I be-PAST you*) to be interpreted?
 - What is the semantics of the copula ‘to be’ in this context?
2. What is the relationship of co-reference between the individuals in the conditional’s antecedent and consequent?
 - By means of which strategy can we trace individuals across worlds?

Before we investigate questions 1 and 2 in order to make advances in quest of the answer to the overarching research question of this thesis, we first need to take into account the (counterfactual) conditional background of *If I were X*-conditionals. The analysis of the meaning of counterfactual conditionality thus constitutes the topic of the next chapter.

3 Theories of Counterfactual Conditionals

When setting out to analyze the meaning of a specific grammatical construction, it is helpful to locate it within a superset of structures about which theoretical knowledge is available. This seems especially fruitful when considering the fact that new insights from subcases might teach us something about the respective general theory, as well. Beyond question, sentences containing the phrase *If I were X* constitute conditional statements; similarly indubitable seems the fact that they make up a subclass of *counterfactual* conditionals. What we would not want to say for this reason is that the overall semantics of counteridenticals differs from that of standard counterfactuals like “If I were ill, I would go to the doctor.” If problems arise in the analysis of the two different readings of *If I were X*-conditionals, we hence conclude that these must stem from a different source. But what exactly does all of this imply? Let us take a step back and be more explicit about what we mean by the notions of ‘conditionality’ and ‘counterfactuality’ before we proceed to analyze the semantics of counteridenticals.⁹

3.1 Definition and Terminology

In the following, I will first define the notions of ‘conditionality’ and ‘counterfactuality’ individually, before I combine them to give a definition of counterfactual conditional sentences.

3.1.1 Conditionality

According to Csipak and Eckardt (2013), conditional sentences can be defined as follows:

A conditional is a sentence that contains two propositions. Usually those propositions stand in some kind of ‘conditional’ relationship to each other (often: one causes the other; if one is true, then we conclude that the other is true as well). Conditionals can, but do not have to, contain the word ‘if.’

In accordance with the given definition, upon hearing a sentence like “If the sun shines, I will go to the park,” one’s intuition about the utterance’s truth-conditions is that the speaker’s going to the park somehow *depends* on whether the sun shines or not. Only if the sun shines, it seems to follow that he will go to the park. In other words: The truth of the clause “The sun shines.” presents a necessary *condition* for the truth of the clause “I will go to the park.” or – more generally speaking – for a sentence of the form “If p, (then) q.” it follows that “[w]hen [p] is true, [q] is also (thereby) made true as well” (Cable 2013).

⁹Parts of this chapter have been adopted from Kauf (2014).

of a line of reasoning which eventually concludes the (possible) truth of the antecedent, i.e. that Jones might indeed have taken arsenic. Even though the reasoning of this structure follows the pattern of the formal fallacy ‘affirming the consequent’ ($p \Rightarrow q; q \therefore p$), which is known to be an invalid form of inference and, thus, challenges the truth of the conclusion, the crucial point here is not that the antecedent might be false – i.e. that there could be other causes for his symptoms besides arsenic – but that it does not necessarily *have to be false*. In short, the statement ‘Hence, he did take arsenic.’ is accepted to follow as a natural consequence from the considerations stated in the preceding sentence and does not stand in contradiction to it, even though it might turn out not to be true. Since counterfactuality can be used in order to argue in favor of the truth of the antecedent of a conditional, we cannot presuppose its falsity (cf. von Fintel 1998: 31).

The second argument is based on the observation that we can also use a counterfactual statement in order to achieve the very opposite effect of the one just discussed, i.e. “to assert the falsity of p without producing redundancy” (Iatridou 2000: 232). The commonly used example to justify this hypothesis was given by Stalnaker (1979) and follows the *modus tollens* form of deduction ($p \Rightarrow q; \neg q \therefore \neg p$), which – in contrast to the inference strategy introduced before – is valid.

- (26) If the butler had done it, he wouldn’t have used an ice-pick. $(p \Rightarrow q)$
 p q
The murderer used an ice-pick. $(\neg q)$
So the murderer must have been someone else. $(\therefore \neg p)$
(cf. von Fintel 1998: 35)

Again, the presupposition of falsity of the first sentence’s antecedent renders an infelicitous result. Unlike the previous example, however, it is not infelicitous because it is contradicted over the course of the conversation; instead, a problem arises as the presupposition of falsity of the antecedent position leads to the redundancy of the last sentence of the utterance. Under such a presupposition, namely, the sentence ‘So the murderer must have been someone else.’ would merely repeat information which has already been assumed, and would not be understood as a conclusion drawn from the combination of the knowledge of the butler’s personality and the evidence of the situation, which is the intended interpretation of this utterance (cf. Iatridou 2000: 232).

All things considered, let us conclude two things from the above discussion: Firstly, counterfactuality will be assumed to be a conversational implicature in all of the following parts of this paper, a position which has been buttressed by the arguments given in the preceding paragraphs. Secondly, the definition of the notions of *conditionality* as well as that of *counterfactuality* can be brought together to schematize the main claim of any theory of counterfactual conditional sentences of the form “If p , (then) q ” in the following way:

- $\{w \mid p(w) = 1\} \subseteq \{w \mid q(w) = 1\}$ \Rightarrow Conditionality restriction
- $w_0 \notin \{w \mid p(w) = 1\}$ \Rightarrow Counterfactuality restriction

3.2 Discussion of Existing Theories

The question that remains open from the given definition of counterfactual conditionals is how we are going to single out the relevant set of p -worlds; or, put differently, how we are going to achieve the non-actuality of its antecedent. A pivotal observation in this regard is provided by the fact that in comparison to indicative conditionals (27-a), counterfactuals contain one/two additional layers of past tense morphology, regardless of the topic's temporal location ((27-b), (27-c)).

- (27)
- a. If Roman *is* at the post office now, he *is* missing the meeting.
 - b. [Roman is not at the post office] If Roman *were* at the post office now, he *would be* missing the meeting.
 - c. [Roman is not at the post office] If Roman *had been* at the post office now, he *would have been* missing the meeting.

(von Stechow 2012: 475)

The two prevalent theories of counterfactual meaning to this day, Iatridou and Ippolito, differ in how they interpret this additional past tense morphology and, consequently, employ different strategies in order to exclude the actual world from the set of antecedent worlds. This section is structured as follows: First, both of the approaches will be sketched individually. Afterwards, it will be discussed whether these theories are able account for both the ambiguity of counterfactuals introduced in the previous chapter as well as the restrictions of validity they have.

3.2.1 Past-as-Unreal Approach – Iatridou (2000)

As the name of the theory suggests, proponents of the past-as-unreal approach do not assign the usual temporal interpretation to the past tense morpheme. Instead, they take the English simple past to introduce a general notion of “distance from reality, non-actuality, or hypotheticality,” which can be applied to different domains (Schulz 2007: 176). One way to formalize this conceptual interpretation of the past tense morpheme has been proposed by Iatridou (2000), who argues that the past tense contributes a skeletal semantics of the form

$$(28) \quad T(x) \text{ excludes } C(x),$$

where $T(x)$ is short for $\text{Topic}(x)$, meaning “the x we are talking about,” and $C(x)$ stands for “the x that for all we know is the x of the speaker” (246). Based on this semantics,

Iatridou has dubbed the English past tense morpheme *exclusion feature* (*ExclF*).

As has been previously stated, the x in schema (28) can be substituted by different domains; more precisely, it can be substituted by the set of times and the set of worlds. When the exclusion feature ranges over the temporal domain, the event is interpreted as real (i.e. actual), since only the evaluation time for the truth of the event is altered whereas the evaluation world remains the same. When, on the other hand, the x varies over possible worlds, an unreal (i.e. counterfactual) interpretation arises. This flexibility of *ExclF* constitutes the strong suit of the past-as-unreal approach as it enables the usual temporal interpretation of the past-tense morpheme, while at the same time being able to do much of the heavy lifting for achieving a counterfactual interpretation of conditional sentences.¹⁰

To see how the exclusion feature works in action, let us consider the two possible cases in turn. First, let the x in (28) vary over times. Under this premise, the following meaning is yielded for *ExclF*:

- (29) “The set of times that we are talking about” excludes “the set of times that for all we know is the set of times of the speaker.”

When adopting Klein’s (1994) terminology, in which the time of the event described in the utterance is called the *topic time* and the ‘time of the speaker’ is referred to as the *utterance time*, i.e. the time at which the utterance takes place, (29) can be paraphrased as

- (30) The ‘topic time’ excludes the ‘utterance time.’

Iatridou considers the future to be modal (cf. 246); hence, the exclusion of the utterance time inevitably leads to a past-tense reading of the morpheme.

Let us turn our attention to the second possible interpretation of *ExclF*, in which the schematic semantics in (28) ranges over the set of possible worlds. In this case, the exclusion feature receives the subsequent interpretation:

- (31) “The set of worlds that we are talking about” excludes “the set of worlds that for all we know is the set of worlds of the speaker.”¹¹
 \Rightarrow *The topic worlds exclude the actual world.*

¹⁰I say ‘much of the heavy lifting’ in this context as it is not always the case that past-tense morphology is needed in order to arrive at a counterfactual interpretation of conditional sentences. This becomes evident when considering the so-called “Indicative counterfactuals” of the type *If you are Santa Clause, I am the Easter Bunny*.

¹¹It should be mentioned that the ‘worlds that for all we know are the worlds of the speaker,’ here referred to as the ‘actual world,’ generate a *set* of worlds, since one’s knowledge of the world is limited to one’s epistemically accessible worlds (cf. Lewis 1986: 27).

As has been stated in the previous section, a counterfactual interpretation of a conditional sentence emerges if the actual world is excluded from the set of possible p -worlds. From the paraphrase given above, it becomes self-evident that this restriction is fulfilled by the past-as-unreal approach.

Iatridou does not explicate which semantics of conditional statements she assumes for her analysis of counterfactual conditionals in addition to the counterfactual flavor the past-tense morpheme adds via *ExclF*. Nevertheless, her line of argumentation suggests that she adopts an analysis along the lines of the standard Lewis-Stalnaker theory of counterfactuals (Stalnaker 1968; Stalnaker and Thomason 1970; Lewis 1973) (cf. Iatridou 2000: 248; Schulz 2007: 177). According to this semantics, a counterfactual ' $p > q$ ' is true at a world w just in case the consequent proposition q holds at all of the antecedent-worlds that are most similar to w . Since she does not consider it crucial for her proposal how the notion of similarity is spelled out, she leaves does not decide for one of the many existing version but rather leaves this as a topic of debate (cf. Iatridou 2000: 248). What is crucial, however, is that a Lewis-Stalnaker-like semantics of conditionals together with her *ExclF* lead to a proposal in which the set of topic worlds emerges as a subset of the p -worlds, i.e. they constitute those counterfactual antecedent worlds which are closest to the actual world with respect to some designated ordering source (cf. *ibid*).

In this thesis, the two main theories of counterfactual conditionals are only to be sketched. The reader is encouraged to consult works like Schulz (2007), von Stechow (2012), or Karawani (2014) for an in-depth discussion of the advantages and disadvantages of the two approaches, therefore. Nevertheless, there are two shortcomings of Iatridou's theory which I do not want to leave unmentioned: Firstly, it is not clear from her proposal when the exclusion feature is predicted to range over times in order to arrive at a temporal interpretation of the past-tense morpheme and when it quantifies over worlds in order to achieve a counterfactual reading. Instead, her theory predicts ambiguity for every sentence with past tense morphology, which, likely, is meant to be disambiguated by means of the context. This imprecision has also been observed by Schulz (2007) or Karawani (2014), who both have tried to formalize Iatridou's approach as a result. The latter, for example, proposes the *Non-Actual Veridicality*-analysis, short NAV, as a refinement. In this analysis, the twofold interpretation of past tense morphology is retained, but the structural position of the morpheme within a sentence is taken to determine which of the two readings arises (see also Karawani and Zeijlstra (2013)). The former aims at disambiguating the meanings by explicating where in the structure the *ExclF* enters the computation.

Schulz's refinement is directly connected to the second shortcoming of Iatridou's theory, i.e. that it lacks precision from a combinatorial semantics point of view. We have seen that it is unclear how the meaning of a sentence is calculated using Iatridou's semantics, which makes it difficult to provide a watertight argument in favor of or against her theory. As has just been stated, one attempt to remedy this deficiency and formalize

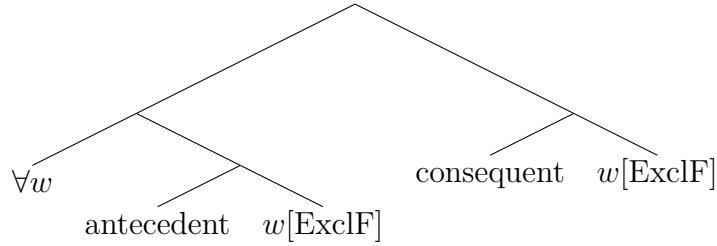


Figure 1: Iatridou’s approach to counterfactuals, represented by Schulz (2007: 177)

the approach has been made by Schulz, who proposes the subsequent logical form as underlying Iatridou’s past-as-unreal analysis of counterfactual conditionals.

The representation is based on Kratzer’s (1986, 2012) *If-clause-as-restrictor* analysis and realizes the past tense morpheme as a feature which is attached to the world arguments of the conditional’s antecedent and consequent. The morpheme’s function in such a structure is to “restrict[] the quantificational domain by excluding the world of the speaker” (Schulz 2007: 177).

3.2.2 Past-as-Past Approach – Ippolito (2013)

In contrast to the past-as-unreal approaches, the past-as-past theories retain the usual interpretation of the past morpheme, i.e. one in which the evaluation time of a sentence is placed prior to the utterance time:

$$(32) \quad \llbracket \text{PAST} \rrbracket = \lambda P_{\langle i, st \rangle}. \exists t' < t: P(t') = 1.$$

In a nutshell, what the proponents of this theory argue is that the extra layer of past in counterfactual conditionals takes us back to a point in time at which the proposition stated in the conditional’s antecedent was still a possibility, even though it might not be so anymore at the utterance time. An obvious implication from this idea is that the further we are taken back in time, the more possible futures open up, as possible worlds are precluded the closer the time argument t converges to the utterance time. One of the most recent and best worked-out of these approaches is Ippolito (2013), which will be sketched in the following.

Simply put, Ippolito’s theory is made up of three main ingredients: The temporal interpretation of the past tense morpheme as given in (32), the branching futures model, which states that for any point in time there exists one, set past but an infinite number of possible futures (Figure 2), and Kratzer’s *If-clause-as-restrictor* analysis, which builds on tenseless ‘bare’ conditionals, i.e. conditionals in which, by default, the verb forms have been replaced by what will be referred to as their ‘adjusted infinitive’ forms. These ‘adjusted infinitives’ look like usual present tense forms, but are assumed to not carry any temporal information. Consequently, the structures’ tense arguments will have to be

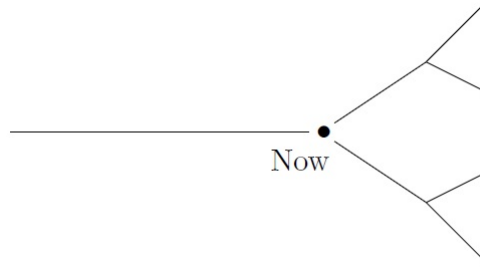


Figure 2: Branching futures model (Karawani 2014: 13)

semantically realized on their own.

The way these ingredients are put together to yield a past-as-past approach is the subsequent: Ippolito (2013) proposes that non-indicative counterfactual conditionals are evaluated with respect to two times, the “accessibility time,” t_a , and the “reference time,” t_r , which can be shifted independently of each other by means of temporal operators (59). She borrows the latter notion of time, the ‘reference time,’ from Reichenbach (1947), who introduces it as the ‘point of reference’ of an utterance, i.e. the time the speaker refers to in his speech act (e.g. When I tried to enter the office (at 7am), the door wouldn’t open. $\Rightarrow t_r = 7\text{am}$) (288).¹² The accessibility time, t_a , on the other hand, is a term coined by Ippolito herself, which she takes to refer to the time of the conditional sentence’s modal operator. As per Ippolito, the different relations between t_a , t_r , and the utterance time, t_c , are what is characteristic of the various types of subjunctive conditionals and have to be captured by the semantic structure in order to give an adequate account of the sentences’ truth-conditions. Consider, for example, the following time relations:

- $t_a < t_r = t_c$
 e.g. “If John were here, the party would be much more fun.”
 \Rightarrow simple past counterfactual conditionals
- $t_a < t_r < t_c$
 e.g. “If John had come, the party would have been much more fun.”
 \Rightarrow past perfect counterfactual conditionals

(cf. Ippolito 2013: 62f)

Ippolito proposes that these time relations are achieved by embedding the ‘bare’ conditional into distinct numbers of layers of past, i.e. one for the simple past counterfactual conditionals and two for their past perfect counterparts (cf. *ibid*). Nonetheless, only one of these past tense parameters is incorporated into the ‘bare’ conditional’s structure, whereas the other, optional one applies to the conditional sentence as a whole. The facultative temporal parameter serves to shift the reference time of the phrase; the first

¹²Note that under Klein’s (1994) analysis, the ‘reference time’ would be referred to as the sentence’s ‘topic time.’ In order to be able to stay as close as possible to Ippolito’s analysis, I am neglecting the consistency of terminology in this part of the thesis and, instead, adopt her nomenclature.

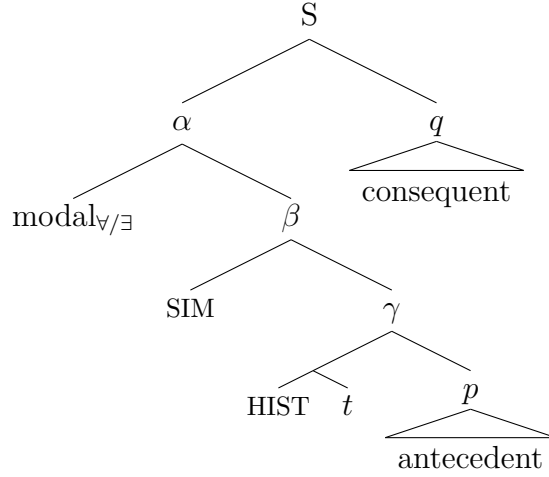


Figure 3: Ippolito's approach to conditionals

layer of past – the temporal morpheme which is realized within the structure – is what is distinctive of counterfactual conditional sentences: it functions to manipulate the tense parameter of Ippolito's accessibility relation, HIST, which she defines as follows:

$$\llbracket \text{HIST} \rrbracket^{c,g,t,w} = \lambda p_{\langle s,t \rangle}. \lambda w'. w' \text{ has the same history as } w \text{ up to } t \text{ and } p(w') = 1. \quad (57)$$

If t is shifted to the past by means of the PAST-operator, the set of worlds about which a claim is made in the consequent only includes those worlds in which the truth of p has been established. More specifically, this means that the actual world is excluded from the set of possible p -worlds via “go[ing] back in the actual world to some time when the antecedent was still not settled and look[ing] at all those ontic alternatives where the antecedent turns out to be true” (Schulz 2007: 167). Hence, Ippolito's theory also successfully excludes the actual world from the set of antecedent worlds and, thereby, fulfills the counterfactuality restriction.

In addition to HIST, which limits the p -worlds quantified over to those historically accessible from the actual world, Ippolito further assumes a stereotypical ordering source, SIM, whose lexical entry is given by:

$$\llbracket \text{SIM} \rrbracket^{c,g,t,w} = \lambda p_{\langle s,t \rangle}. \lambda w'. p(w') = 1 \wedge \neg \exists w'' [p(w'') = 1 \wedge w'' <_w w'] \quad (58)$$

This similarity function ranks p -worlds according to how close they are to the actual world w under the condition that p is true and ensures that a claim is made only about the most similar of these p -worlds – it neglects all remote or somehow ‘strange’ antecedent worlds. As represented in Figure 3, both HIST and SIM can be represented overtly in Ippolito's underlying structure (cf. 57ff, Schulz 2007: 169).

Nevertheless, the structure can be simplified by means of the introduction of the modal WOLL, which merges the conditional's modal force, accessibility relation and

ordering source in order to arrive at the following skeletal semantics for conditional sentences. The only inputs still needed in this structure are the time argument t that manipulates HIST as well as the antecedent and consequent propositions, p and q :

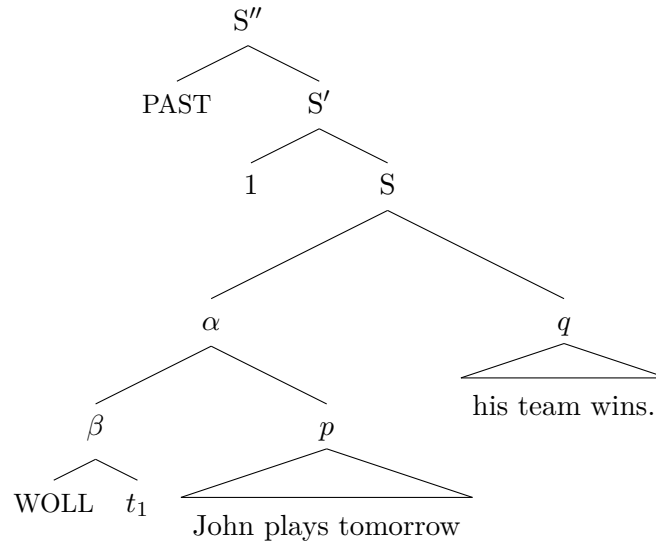
$$\llbracket \text{WOLL} \rrbracket^{c,g,t,w} = \lambda t'_i. \lambda p_{\langle s,t \rangle}. \lambda q_{\langle s,t \rangle}. \forall w' \in W [w' \in \text{SIM}_w (\text{HIST}_{w,t}(p)) \longrightarrow w' \in q].$$

(cf. Ippolito 2013: 60)

In order to see how Ippolito's theory works in contrast to Iatridou's, consider the following situation: John's team has a soccer match tomorrow. John is the best player of the team but has broken his leg, so he cannot play. If he could play, his team would have a good chance of winning, but without him, they are likely to lose. In such a situation, John's mom could say to one of her friends:

(33) If John played tomorrow, his team would win.

As per Ippolito, the following logical form belongs to sentence (33).



(Ippolito 2013: 60)

The reference time is the utterance time, therefore, only one layer of past – the one which turns the sentence contrary-to-fact – is needed. As expected from the preceding explanation, this layer of past is applied within the ‘bare’ structure of the conditional and is immediately merged with the operator WOLL. This way, it can access the temporal parameter of its historical accessibility relation. WOLL then takes the antecedent and consequent propositions as its inputs, before the temporal argument is eventually evaluated by means of the λ -Abstraction Principle to yield the truth-conditions of the conditional structure. When calculating the truth-conditions of (33) via this strategy, one obtains the following result:

$\llbracket (33) \rrbracket = 1$ iff $\exists t' < t: \forall w' \in W [w' \in \text{SIM}_w (\text{HIST}_{w,t'}(\lambda w''. \text{John plays tomorrow in } w'')) \rightarrow w' \in \{w''' \mid [\lambda w''. \text{his teams wins in } w''] (w''') = 1\}]$
 \Rightarrow *Sentence (33) is true if and only if “there is a time t' before the utterance time t_c such that all the worlds historically accessible at t' from the evaluation world w where John will play tomorrow and that are (overall) closest to w are worlds where his team will win.”*

(cf. Ippolito 2013: 61)

Ippolito’s model is more complex than the sketch given above; it also includes restraints on the presuppositions of the antecedent and consequent clauses, which must be satisfied at worlds which are historically accessible from both the accessibility time and the reference time, in order for the model to circumvent the trivial truth of counterfactual conditionals (e.g. the existence presupposition for individuals) (cf. 62ff). Even though the importance of these refinements of Ippolito’s theory are acknowledged, for the purpose of this paper an understanding of the reduced version introduced above suffices – the consideration of presuppositions will not turn out to be the crucial factor in favor of, or against, the model’s ability to capture the meaning of *If I were X*-conditionals.

As with the Iatridou’s past-as-unreal approach, Ippolito’s theory has its strong suits as well as shortcomings, which I will not be able to discuss in detail in this work. One of the proposed disadvantages of her theory, for example, consists in its dependence on the historical accessibility relation, HIST, which is not always philosophically trivial (cf. e.g. von Stechow 2000: 15, Karawani 2014: 476, Kauf 2014: ch. 4.4). Nevertheless, for this thesis – and in contrast to Iatridou – the strong compositional orientation of Ippolito’s theory as well as its precision is deemed an important advantage.

3.2.3 Application to Counteridenticals

Previously, the two most prominent theoretical approaches to the meaning of counterfactual conditionals have been sketched and it has been shown how both successfully employ different strategies to exclude the actual world from the set of antecedent worlds in order to enable a counterfactual interpretation of conditional sentences. Knowing that both theories, thus, render equivalent truth-conditions, it suffices to evaluate one’s sample sentences on the basis of only one of the analyses when testing a hypothesis about counterfactual meaning. As both proposals have been argued to have disadvantages as well as advantages in comparison to the other, there is none which can objectively be considered superior. Based on my purely subjective choice, therefore, Ippolito (2013) has been selected as the test theory for this thesis, since its careful dissection of the conditional semantics is deemed to facilitate the analysis.

Section 3.1 established that *If I were X*-conditionals constitute a subset of counterfactual conditional sentences. Given the findings from sections 3.2.1 and 3.2.2, the

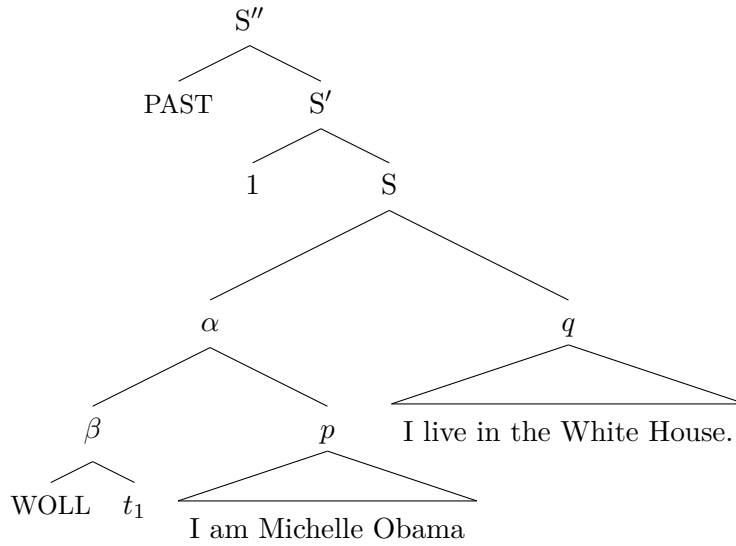
zero hypothesis which emerges with regard to their analysis is that both of the discussed theories should be able to capture their intended meaning without any problems. Let us test this hypothesis on the basis of Ippolito’s proposal.

‘Imagine’ Counteridenticals I

As a first trial sentence, reconsider sentence (23-a), restated here as (34), where the intended interpretation of the antecedent *If I were X* is that of an ‘imagine’ reading.

(34) If I were Michelle Obama, I_{as M.O.}’d live in the White House.

As before, the according logical form is given by



and, via employing the same strategy as described before, the sentence’s truth-conditions amount to

$\llbracket (34) \rrbracket = 1$ iff $\exists t' < t: \forall w' \in W [w' \in \text{SIM}_w (\text{HIST}_{w,t'}(\lambda w''. \text{I am Michelle Obama in } w'')) \rightarrow w' \in \{w''' \mid [\lambda w''. \text{I live in the White House in } w''] (w''') = 1\}]$
 \Rightarrow *Sentence (34) is true if and only if there is a time t' before the utterance time t such that all the worlds historically accessible at t' from the evaluation world w where I am Michelle Obama and that are (overall) closest to w [according to SIM] are worlds where I live in the White House.*

From this semantics, we deduce that the truth of the entire clause crucially depends on the interpretation that we ascribe to the identity statement “I am Michelle Obama.” Here is what we know: I am not Michelle Obama. Nevertheless, Ippolito suggests that were we to go back in time to a moment where the antecedent was still unsettled, e.g. to a point in time at which both Michelle Obama and I had not been born yet, and were we to look at all those worlds in which the antecedent proposition turns out to be true, i.e.

those worlds in which I am the same person as Michelle Obama, then the closest p -worlds according to a stereotypical ordering source would form a subset of the set of worlds in which I live in the White House. Let us give thought to the possibility that I could also live anywhere else and *a prima facie* the antecedent could still be true; however, if I turned out exactly the way Michelle Obama has, or at least possessed all of her contextually relevant properties, as is implied by the asymmetric antecedent *If I were Michelle Obama* – whose meaning vitally differs from that of *If Michelle Obama were me*¹³ –, then I would be married to the president of the US and, as a result, I would indeed live in the White House at the utterance time; all other worlds would be further away from the actual one according to SIM under the truth of p .

Given this line of argumentation, at a first glance, one might think that Ippolito's theory is able to capture the sentence's correct meaning under a specific premise regarding the identity relation. Nonetheless, her assumption of a past which is common to both the actual world as well as a counterfactual p -world is not philosophically trivial for counteridenticals. In essence, Ippolito thereby implies that there exists a point in the past up to which all of the facts of the world were the same for the entire set of possible worlds, and after which the worlds branched and developed into different continuations, i.e. in some of those possible worlds I ended up being myself and in other possible worlds, I ended up being Michelle Obama. It seems dubitable, however, that such a point in time should exist.

Let us assume that we go back in time to January 16, 1964, the day before Michelle Obama was born and when I was still long from being planned, and let us further assume that whatever ghost happened to develop into me in the actual world developed to be her in those possible worlds. Given SIM, in such p -worlds, my parents' lives would not be affected by this identity swap and, as a result, could develop in the same way as they have in the actual world. In particular, my parents could have a child on the same day that I was born. Since prior to this child's birth, I would already have developed into Michelle Obama, however, it seems implausible to think that the child could be me. The question which immediately presents itself as a result is how to even define 'myself' in such a scenario: Under the most extreme interpretation of the identity relation *I am Michelle Obama*, 'I' end up having all of Michelle Obama's properties and none of myself; hence, it does not seem expedient to think of *me* as Michelle Obama in such worlds since it is technically only her that exists. At the same time, my parent's child, who will have all of the same properties that I have in the actual world, cannot be considered to be me for obvious reasons, either. In conclusion, both of the individuals that would be feasible candidates for the referent of the pronoun *I* in the conditional's consequent can be excluded via logical reasoning.

¹³The different meanings of identity relations will be discussed more thoroughly in the next part of this thesis. Also see Kocurek (to appear) for a discussion and additional examples of this contrast.

I have argued that we have reason to be skeptical about Ippolito’s historical accessibility relation in the context of counteridenticals. Nevertheless, for the sake of the analysis and the time being, let us assume its validity and let us test how Ippolito’s theory deals with the second application of *If I were X-counterfactuals*, i.e. its ‘advice’ use, under this premise.

‘Advice’ Counteridenticals

To the end of analyzing ‘advice’ counteridenticals, consider sentence (1-a), restated here as (35).

(35) If I were you, I’d buy the blue dress.

Since the only thing which is different from the previous sentence is the content of the antecedent/consequent propositions but not the tense schema, Ippolito’s theory proposes the same underlying logical form as before, as well as a parallel analysis. Consequently, the truth-conditions of (35) are as follows:

$$\llbracket (35) \rrbracket = 1 \text{ iff } \exists t' < t: \forall w' \in W [w' \in \text{SIM}_w (\text{HIST}_{w,t'}(\lambda w''. \text{I am you in } w'')) \longrightarrow w' \in \{w''' \mid [\lambda w''. \text{I buy the blue dress in } w''] (w''') = 1\}]$$

\Rightarrow *Sentence (35) is true if and only if there is a time t' before the utterance time t such that all the worlds historically accessible at t' from the evaluation world w where I am you and that are (overall) closest to w [according to SIM] are worlds where I buy the blue dress.*

From this paraphrase we realize that ‘advice’ counteridenticals, and thus counteridenticals in general, present a challenge to existing semantic theories of counterfactual meaning and that our zero hypothesis might have to be rejected as a result. This insight is grounded in the fact that the counteridentical’s analysis under the same premise regarding the meaning of identity statements as successfully assumed above does not give the correct semantics for (35): Were the presupposition correct and we would consider those alternative worlds at which the antecedent is true, i.e. I am you (meaning that I possess all of your (contextually relevant) properties), and – again, crucially – we would not be the same person because you would be me, then the ordering of those p -worlds according to SIM would ensure that the most similar antecedent world to the actual world is one at which the subject of the consequent clause, $I_{\text{as } A}$, necessarily does whatever the addressee would do in the same situation of dress shopping in the actual world, i.e. $I_{\text{as } A}$ would choose the same dress as the addressee would in the actual world (cf. e.g. Gordon 1995: 740, Williams 1973). Given that we are in a situation in which the speaker wants to give the addressee a piece of advice, however, this does not seem to be the result we desire.

Let us take a step back and rephrase this observation to make it more explicit: Assume that there exists a point in time up to which all possible worlds share a past and at which the worlds start to branch into ones at which a) I develop into the actual me and b) I become the same person as you, at least with respect to the contextually salient properties for the conversation (i.e. here: taste in fashion). According to Ippolito’s theory of counterfactuals, we only consider the latter set of possible worlds, the set of p -worlds, for the analysis, which is ordered with respect to their overall stereotypical similarity to the actual world in a next step. The truth of the consequent clause is then evaluated at the closest of these worlds. If, however, the subject of the consequent is the same as the addressee, at least with respect to his taste in fashion, and if except for the truth of this counterfactual proposition, the counterfactual worlds remain as close to the actual one as possible, i.e. in particular the choice of dresses under discussion remains the same, etc., then this individual, possessing the addressee’s taste in dresses in that world, will necessarily choose the same dress as the addressee would in the same situation in the actual world. Whereas there might be situations in which the speaker will advice something that the object of the counterfactual’s antecedent clause would have done without this piece of advice, too (e.g. there may exist situations in which the addressee in (35) would want to buy the blue dress anyways, independently of the speaker’s advice), this cannot be taken as the default situation for the application of ‘advice’ counterfactuals.

In situations in which the contextually relevant properties of the advisor and the advisee differ, it seems self-explanatory, therefore, that the result we have obtained is not the one we desire, since we have argued before that in ‘advice’-counterfactuals, it should be the properties of the *speaker* which determines the consequent proposition and not that of the person he is trying to give a piece of advice to. Pelletier (2004) phrases this observation in the following, fitting way: “If we identified with the other [=the object of the antecedent clause in a counterfactual] [in the contextually relevant properties], we then would lose the very advantages that makes our advice worthwhile: the special know-how or the independent judgment we are supposed to have” (148). Complete identification with the object of the antecedent clause (regarding the contextually relevant properties) prevents us from being able to give them practical advice, therefore.

The fact that this contrast between the ‘imagine’ and the ‘advice’ reading is not trivially predicted by standard theories of counterfactuals becomes even more evident when looking at the *technical* challenge the ‘advice’ reading brings about in contrast to the structure’s ‘imagine’ usage: They incorporate two instead of the commonly-assumed one dissimilarity with respect to the actual world. First of all – as in standard counterfactual sentences – the antecedent proposition is conveyed to be contrary-to-fact (in (35): I am not you in the actual world, $w_0 \notin \{w \mid p(w) = 1\}$). Additionally, however, we have just seen that for ‘advice’ counterfactuals it is further the case that not all p -worlds must be

q -worlds – a property which stands in marked contrast to the definition of counterfactual conditionals given in section 3.1. In concrete terms, this means that under the meaning of identity assumed so far, the additional dissimilarity in sentence (35) manifests itself in the possibility of it not being the case that in all those worlds in which the speaker is assumed to be the same person as the addressee, he buys the same dress as the addressee normally would. In other words: There may exist worlds at which the speaker is the identical to the addressee with respect to the contextually relevant properties, but at which this property-modified speaker still does not behave in the same way the addressee would. In a nutshell this means that the conditionality restriction ($\{w \mid p(w) = 1\} \subseteq \{w \mid q(w) = 1\}$) is not fulfilled in such situations.

A convincing argument in favor of this observation is provided by the fact that we can even apply ‘advice’ counteridenticals in situations where *no* p -world is a q -world under the premise of identity statements adduced so far. To illustrate this, consider the following example:

- (36) CONTEXT: You used to own a beautiful black dress, but you forgot it in your hotel room in Rome. I don’t know that you’ve lost the dress. Since you have a date tomorrow night, you’re asking me what to wear. I say:
- a. If I were you, I’d wear that beautiful black dress.

Given that the addressee no longer owns the dress the speaker is talking about, the following semantics seems at odds with our intuition:

Sentence (36) is true if and only if for all worlds at which the speaker is the same as the addressee (with respect to the contextually relevant properties), and all other facts of the world are as close to those in the actual world as possible, i.e. in particular, the black dress is still somewhere in Rome, the subject of the consequent clause, $I_{as\ A}$, being contextually similar to the addressee, wears this dress to the date.

All things considered, we conclude that even though the counterfactuality restriction is fulfilled in (35) and (36), under Ippolito’s theory, conditionality is inherently rejected for ‘advice’ counteridenticals:

- $w_0 \notin \{w \mid p(w) = 1\}$ \Rightarrow Counterfactuality restriction \checkmark
- $\{w \mid p(w) = 1\} \subseteq \{w \mid q(w) = 1\}$ \Rightarrow Conditionality restriction \times

The starting point of our analysis was provided by the observation that counteridenticals should have the same overall semantics as standard counterfactuals like “If I were ill, I would go to the doctor.” Since these are known to fulfill both of the restrictions listed above, we infer that the semantics we have obtained for example (35) cannot be regarded a satisfactory result. It seems that the problem can be solved if one introduced appropriate

conditionality conditions; Nevertheless, fixing the conditionality of ‘advice’ counteridenticals turns out not to be a trivial matter: The previous discussion suggests that only an identity relation which ensures a one-hundred percent (contextually relevant) conformity of the antecedent clause’s subject with its object leads to conditionality effects between the p - and q -worlds in a counteridentical. At the same time, it seems self-explanatory that such an absolute identity relation inevitably brings about an ‘imagine’ reading of the antecedent *If I were X*. To put things in a nutshell, we realize that it is impossible to repair the standard semantics of counterfactual conditionals with regard to counteridenticals while still keeping their ‘advice’ reading feasible under the meaning of identity statements assumed so far.

‘Imagine’ Counteridenticals II

Up to this point, we have argued that if we neglect the philosophical problems which arise with respect to Ippolito’s historical accessibility relation and if we further assume a specific way of interpreting the identity relation in the counterfactual’s antecedent, then it seems that the structure’s ‘imagine’ uses can be accounted for through standard theories of counterfactual meaning. By contrast, ‘advice’ counteridenticals have been shown to challenge the validity of these theories under the same assumptions. In the following, it will be argued that even under the premises which have kept the ‘imagine’ reading of counteridenticals feasible so far, this reading, too, can be problematic for existing theories. In order to buttress this claim, let us consult another one of the previously discussed examples.

(37) If I were you, I’d feed my cat every day.

$I_{as A} \quad I_S$

Before, it was left to the reader to think of a possible situation in which the sentence with this interpretation can be used. In order to prove that this reading indeed exists, let us now consider the following scenario: You are in a bad mood and I am trying to cheer you up. My strategy to achieve this is by listing all the things which you manage to do in your everyday life, which includes feeding my own cat every day, a task you have taken over because I tend to come back home from work rather late. In such a situation, I could say something like “You manage to do so much every day, you should be proud. See, *if I were you, I’d work ten hours every day, I would go to the gym and I would still find time to go out with my friends, do chores and even feed my cat every day*. See, you even manage to do some of my chores while all I manage to do is work - how do you do that?”

But why does sentence (37) now present a challenge to the existing semantic theories? The reasoning follows along the same lines as above even though the problem is a different

one. In this example, it is not the conditionality restriction that is violated, as all the worlds in which I am (contextually relevantly) similar to you are also worlds in which I, as you, do the feeding. Here, the problem arises due to the second indexical in the consequent clause, which, following strong intuition, refers to a different *I* than the first one – possibly to the speaker of the utterance. If the antecedent takes us to worlds in which I am you and which are otherwise closest to w_0 with regard to SIM, we have seen that the speaker-*I* should not even exist anymore and, as a consequence, should not be able to be referred to in the consequent, either. The interpretation which Ippolito’s theory would yield in this context is the following

(38) If I were you, I’d feed my cat every day,

$$I_{as A} \quad I_{as A}$$

where only one *I*, the speaker who has turned into the same person as the object of the antecedent clause, exists.

In conclusion, we infer that Ippolito’s theory is not trivially able to account for the different readings of counteridenticals for the following reasons: Firstly, her assumption of a past which is common to the actual as well as the counterfactual *p*-worlds is not philosophically trivial. Secondly, under the assumption of a concept of identity suitable for the ‘imagine’ reading of *If I were X*-conditionals, their ‘advice’ reading cannot be accounted for as it incorporates two dissimilarities with respect to the actual world instead of the one predicted by Ippolito’s theory. Lastly, the proposal is not trivially able to account for multiple referents of the first person pronouns which can occur in the consequent clause of the counteridentical.

Comparison with Iatridou (2000)

In the beginning of this chapter, it has been explained why we do not expect a different result with respect to our hypothesis from Iatridou’s theory than we have received from Ippolito’s proposal; hence, it will not be discussed here in detail. Nonetheless, not all of the problems which Ippolito’s analysis runs into also present a challenge for Iatridou’s theory and vice versa. Especially striking in this regard seems the observation that Iatridou does not assume a historical accessibility relation for her proposal, which means that the philosophical problem of identity discussed above does not emerge from her approach. As a result one might conclude that Iatridou’s theory is more suitable to deal with the ‘imagine’ reading of counteridenticals, in which the speaker is identifying himself with the object of the antecedent clause to a much vaster degree than in advice counterfactuals, up to full identity.

Nevertheless, there seems to be a tradeoff at stake: The only reason why Iatridou’s theory does not run into a similar problem as Ippolito’s seems to be that she does not

specify any restrictions of accessibility on the contrary-to-fact worlds, or, more generally speaking, opts not to spell out specifically how the semantics of a conditional under her proposal are to be calculated (cf. section 3.2.1). This imprecision, however, raises the question of how p -worlds are accessed and, as an immediate result, of which worlds should be considered the closest antecedent worlds to the actual world. We know that if the x of Iatridou’s *ExclF* quantifies over the domain of possible worlds, w_0 is excluded from the set of possible p -worlds, which is then ordered with respect to the elements’ similarity to the actual world. Since it is unclear which factors of similarity are the decisive ones, however, her analysis appears to be much more context-dependent than Ippolito’s, which may lead to independent difficulties. Hence, even though it seems easier to rule in challenging examples under a weaker proposal of counteridenticals, we see that their lack of precision leads to problems with respect to different facets of the analysis.

What is more, even if we concluded that Iatridou’s theory was superior to Ippolito’s for the analysis of imagine counteridenticals, we would quickly realize that it does not perform any better with respect to advice counterfactuals. Although Iatridou does not explicate the conditional semantics underlying her proposal, there is a strong intuition that she assumes a Lewis-Stalnaker-like analysis in which “quantification over possible worlds is explicit” (Schulz 2007: 177). If, however, Iatridou’s *ExclF*’s only function is to exclude the actual world from the set of antecedent worlds and if the topic worlds thus determined are subject to the conditionality restriction ($\{w \mid p(w) = 1\} \subseteq \{w \mid q(w) = 1\}$), the second dissimilarity advice counterfactuals incorporate with respect to the actual world cannot be trivially accounted for by Iatridou’s theory, either. Similar to Ippolito (2013), no degree of freedom enabling the q -worlds in an advice counteridentical not to be p -worlds under the trivial hypothesis of identity assumed above is part of her proposal. It will be left to the readers to convince themselves of this line of argumentation by means of the example sentences discussed for Ippolito’s theory in this section.

3.3 Concluding Remarks

We have seen that standard theories of counterfactual meaning are not trivially able to capture the correct semantics of counteridenticals in general, and those of sentences like (35) and (37) in particular. But if these sentences’ semantics are unclear, how is it that we can still utter them felicitously and interpret them in the intended way?

From the preceding discussion it has become evident that the problem of analyzing the meaning of counteridenticals does not arise with regard to the theory of counterfactual conditionality per se, but that it rather stems from the theory’s union with another semantic problem: that of the meaning of identity statements. This becomes particularly evident when remembering that as early as in the very first application of Ippolito’s theory to counteridenticals, we have stated that ‘the truth of the entire [conditional]

clause crucially depends on the interpretation that we ascribe to the identity statement [in its antecedent]’ (see the discussion of example (34)). Without thinking about the technicalities, we assumed two properties of identity statements which seemed plausible for the analysis of the ‘imagine’ use of *If I were X*-conditionals: Firstly, we inferred from the data that the ‘imagine’ interpretation of the identity statement requires complete identification of its subject and its object. On this note, however, we argued that such a ‘complete’ identification does not need to hold globally but that it can be restricted to the contextually relevant properties of the situation. Secondly, we realized that this identification process only works in a specific direction, meaning that it is not symmetric: In order for an ‘imagine’ interpretation of a counterfactual to arise, it is crucial that the properties determining the truth of the consequent clause are those of the object of the antecedent, at least in the default case where no designated prosody singles out a specific entity as the property provider. These assumptions are in line with our typological discussion of counterfactuals in section 2.

In addition to the asymmetric (at least property-restricted) complete identification of the subject and the object which we assigned to identity statements in order to make the standard analyses work for the ‘imagine’ reading of counterfactuals, we saw that the counterfactuals’ ‘advice’ use requires yet another interpretation of the identity statement under the same assumptions for the semantics of counterfactuals. In particular, we found out that the identification process between the speaker and the object of the antecedent clause must not be complete, since otherwise, an ‘imagine’ reading emerges instead of an ‘advice’ one. If we wanted to keep the assumed semantics of counterfactuals and identity statements, we would have to compensate this difference by another means, therefore.

It has been pointed out to me by von Stechow (p.c.) that the similarity relation might be a good candidate to help us achieve the desired compensation. Even though much evidence points into the direction that the problem counterfactuals pose is more relevant to the meaning of identity statements than it is to the semantics of counterfactuals, let us briefly explore this possibility. Von Stechow’s suggestion to exploit the flexibility of the similarity relation for the purpose of accounting for a difference in interpretation of the same counterfactual is in line with Quine (1950, 1960), who observed that the similarity relation in a counterfactual is not well-defined but incorporates a degree of freedom regarding the factors which determine the hypothetical *p*-worlds’ ‘overall similarity’ to the actual one (cf. *ibid*). As per Quine, the speaker’s intended message as well as the context of utterance alone decide about the truth of a counterfactual statement, which is why pragmatically incompatible sentences can be true in different contexts. To illustrate this, consider Quine’s famous pair of conditionals:

- (39) a. If Caesar had been in command [in Korea], he would have used the atom bomb.

- b. If Caesar had been in command [in Korea], he would have used catapults.
(Quine 1960: 222)

Whereas in (39-a), Caesar’s ruthlessness is understood to be the crucial determiner of similarity and the facts of history, i.e. which weapons Caesar would have actually used to lead a war in his lifetime, are compromised to this end, in (39-b) the speaker focuses on keeping Caesar unmodernized and, thus, the technologies of warfare rigid; here, the technological process up to the 1950s is neglected in ordering the set of possible p -worlds (cf. Lewis 1973: 67). In short, Quine argues that a counterfactual is ambiguous: It is silent about “what traits of the real world to suppose preserved in the feigned world of the contrary-to-fact antecedent” (Quine 1960: 230), which means that the hearer must infer them from a combination of the consequent clause’s content and the context of utterance, or – as Quine puts it – guess them “from a sympathetic sense of the fabulist’s likely purpose in spinning his fable” (ibid.) (cf. Johnson-Laird 1983: 60).

With respect to the research question of this thesis, we derive that we might be able to remedy the problems we have encountered in the analysis of counteridenticals by adapting the similarity relation in such a manner that for each occurrence, it selects for exactly those p -worlds in which the identity statement is interpreted in the way demanded by the context and intended by the speaker. Nonetheless, this seems like not only a rather vague solution but also a tedious exercise as we would have to determine the decisive factors of similarity for each context individually. What is more, the alteration of the similarity relation – a component which Ippolito’s and Iatridou’s theories both crucially depend on – might complicate the proposals or even introduce ambiguity. This seems undesirable since both approaches make the correct predictions for many sentences. Lastly, we realize that even if we were to choose this strategy, we would still have to figure out the intended meaning of the identity statement for each context in order to adapt the decisive factors of similarity accordingly; hence, this strategy likely leaves us with more work to do than if we were to solely interpret and adapt the meaning of the identity statements in the first place. As a result, I have chosen not to pursue this strategy but to stay with my initial approach, i.e. adapting the meaning of identity statements to the different usages of *If I were X*-conditionals.

4 The Semantics of Identity Statements

The previous discussion has shown that exploring the range of meanings of identity statements and incorporating them into existing theories of counterfactual meaning seems to be the preferred strategy to capture the semantics of counteridenticals up to this point. Before we are able to *adapt* the semantics of identity statements in this context, however, we need to learn not only what they mean individually but also where exactly in the analysis of *If I were X*-conditionals they contribute to the meaning. To this end, let us first start by giving a general definition of the notion ‘identity statements’ and an overview of the terminology usually used to capture this notion formally.

4.1 Definition and Terminology

The *Blackwell Dictionary of Western Philosophy* defines the term ‘identity statement’ as follows:

A statement in which an expression of **identity**, such as “is” or “are,” unites two expressions aiming to identify the same thing or kind of thing. All identity statements are symmetric, that is, if “A is B,” then “B is A.”

At the bottom of such a notion of identity statements is an equative interpretation of the copula ‘to be,’ which has traditionally been represented by the symbol ‘=’ in the literature (e.g. $\llbracket \text{Mount Everest is Chomolungma} \rrbracket = 1 \iff [\llbracket \text{Mount Everest} \rrbracket = \llbracket \text{Chomolungma} \rrbracket] = 1$).

Equative copular clauses stand in contrast to predicational ones, where a property of the subject referent is predicated, e.g. *The hat is big./Rabbits are cuddly*. In such contexts, no equality of the expressions identified with each other by means of the copula is assumed, meaning that the parses ‘ $\llbracket \text{the hat} \rrbracket = \llbracket \text{big} \rrbracket$ ’ and ‘ $\llbracket \text{rabbit} \rrbracket = \llbracket \text{cuddly} \rrbracket$ ’ are not intended (cf. Mikkelsen 2001, based on Higgins 1979: 204ff). This insight is in line with the restriction the above definition poses on identity statements, i.e. that they equate only ‘the same thing or kind of thing.’ Instead of implying equality of entities, we understand predicational copular clauses to connect two objects by means of a set membership- or a subset relation (i.e. $\llbracket \text{the hat} \rrbracket \in \llbracket \text{big} \rrbracket / \llbracket \text{rabbit} \rrbracket \subseteq \llbracket \text{cuddly} \rrbracket$).

Since this chapter is dedicated to the meaning of identity statements, the predicational use of the copula appears to be of no importance. Nevertheless, it will turn out to be the case that the distinction between the two kinds of copular clauses is not as well-defined as one would assume from the above definition. When differentiating the notion of identity statements and analyzing its semantics in the next section, let us keep the two possible relations the copula ‘to be’ induces in mind, therefore.

4.2 Problematization and Differentiation of the Notion ‘Identity Statement’

Now that a general idea of what the term ‘identity statement’ means has been developed and how it is connected to the different interpretations of the copula ‘to be,’ I will proceed to problematize and differentiate it in light of the usages relevant to the analysis of counteridenticals. To make out these different usages, let us briefly reconsider the problems that we have encountered in the previous chapter.

The most obvious way in which the meaning of identity statements contributes to the meaning of *If I were X*-conditionals is via the relation they set up between the subject and the object of the conditionals’ antecedents. Here, an overt identity predicate, “were,” identifies two expressions of the same kind, i.e. individuals, with each other, which, in accordance with the above definition, should yield a symmetric identity relation representable by the statement $[[I]] = [[X]]$. Nevertheless, the previous chapter has shown that the notion of identity needed to successfully capture the semantics of counteridenticals cannot be this trivial. Without yet knowing about the technicalities connected to identity predicates, we inferred that the subject and the object in the antecedent of counteridenticals are not related in a symmetric way, but that they are identified with each other by means of a unilateral, asymmetric, ‘complete’ identity relation in the case of the ‘imagine’ interpretation of counteridenticals, and a unilateral, asymmetric, partial identity relation in their ‘advice’ use. As a result, we conclude that the standard definition of identity statements/predicates given above needs to be extended by an asymmetric usage of the term in certain contexts, as well.

In addition to this explicit meaning contribution, also the second problem we have encountered in the previous chapter can be traced back to an application of identity statements: the ambiguity of first person pronoun reference in the structure’s consequent. Even though in this case no identity statement overtly appears in the counteridenticals’ surface representation, the existence of covert identity relations in the analysis of counteridenticals can be derived from the different possibilities of interpreting first person pronouns, which calls for a means to relate each occurrence of them to a designated individual in a symmetric way. Since the previous chapters have shown that some of these relations seem to hold across worlds, we infer that the given definition of identity statements needs to be refined to incorporate a notion of a world-sensitivity, as well.

Before the analysis of the different usages of identity statements in counteridenticals stated above, let us take a step back and properly buttress our intuition that in speech acts which are about more than one world, identity relations cannot be interpreted in the trivial way assumed by the above definition. For this purpose, we adduce independent arguments from the (historical) literature on attitude contexts – a topic which has been much more discussed than *If I were X*-conditionals in this context.

4.2.1 Identity Statements in Cross-World Contexts

Let us recall once more that from definition of identity statements given above we infer that *a prima facie* identity relations should be interpreted in the same way as ‘equality.’ This intuition is in line with the pioneering work on the analysis of identity statements in natural language, Frege (1982), who argues that all kinds of identity statements can be represented by the formula ‘ $a = b$,’ a representation which implies that the expressions $\llbracket a \text{ is } b \rrbracket$ and $\llbracket b \text{ is } a \rrbracket$ have the same meaning (cf. Percus and Sharvit 2014).

Nevertheless, it was also already Frege who discovered that the equivalence of two rigid designators does not always constitute a necessary truth, since the symmetry restriction it incorporates can be circumvented under attitude contexts. This insight he deduced from his realization that in spite of their parallel representations identity statements can have different epistemic values: Some of them hold necessarily/*a-priori* and, thus, possess only a marginal epistemic value (e.g. Cicero = Cicero), whereas others are learned to be true *a-posteriori* and are regarded as epistemically valuable, therefore (e.g. Cicero = Tully). On the basis of this observation, Frege discussed whether the two-place equality relation ‘=’ in identity statements should hold between names or between their referents in the world, which leads us to the conclusion that the equivalence of objects is not always “*epistemically* necessary, even though it seems *alethically* and *temporally* necessary,” where ‘alethical rigidity’ implies cross-world designation of the same individual (LaPorte 2016: 1.4). As a result, we accept pairs of sentences like the following to be compatible with each other:

- (40) a. Cicero is Tully.
 b. John believes that Cicero and Tully are two different persons.

A related insight which is usually also accredited to Frege, is that for identity statements embedded in attitude contexts, the *Principle of Identity Substitution* fails. This principle states that in any true statement, we may substitute a term by a true equivalent thereof and the resulting statement will also be true ($a = b$; $p(a) \therefore p(b)$) (cf. Quine 1943: 113). The failure of this principle in the context of attitude statements similarly becomes evident when considering examples like the following:

- (41) Cicero is Tully. $(a = b)$
 John believes that Cicero is a great philosopher. $(p(a))$
 John believes that Tully is a great philosopher. $(\therefore p(b)) \not\vdash$

Clearly, the deduction does not go through, since it might be the case that the concept John has of Tully is completely different to the one he holds with respect to Cicero; in fact, he might not even be acquainted with the name ‘Tully’ and, therefore, could not possibly associate it with the philosopher. Even though this inference is not valid on an

epistemic level and the Principle of Identity Substitution fails in all of John's deontic worlds, it is interesting to think about the fact that on a metaphysical level, the principle remains solid: Knowing all of the relevant facts of the the world John lives and holds beliefs in, namely, one could infer that the referent of the name 'Cicero' is in fact equal to the one of 'Tully' (cf. Quine 1943: 114).

Up to this point, we have only considered examples of identity statements, where the identity is either both epistemically and metaphysically true (i.e. Cicero = Cicero) or metaphysically true but does not hold in all epistemic/hypothetical worlds (i.e. Cicero = Tully; I = the speaker). Nevertheless, let us give thought to the idea that there also exist usages of identity statements, where the identity relation holds epistemically even though it is not alethically and temporally true. These usages seem particularly interesting for the analysis of *If I were X*-conditionals, since their consequents make a claim about worlds in which two inherently incompatible entities have been identified with each other by means of the antecedent clause (cf. Declerck and Reed 2001: 100). We will see that there exist two possible usages of such identity relations, one which is symmetric and one which is asymmetric in one's epistemic worlds. Both cases will turn out to be relevant for the analysis of counteridenticals: the former one with regard to pronoun resolution and the latter one with regard to the relationship the antecedent clause sets up between the individuals under discussion. Since we will first dedicate our attention to the case of pronoun resolution, let us reserve the discussion of the asymmetric use for later.

Examples of symmetric equality relations which are only epistemically but not metaphysically true arise in situations in which two distinct entities in the actual world merge into one single entity in our attitude worlds. On a related note, we can, for example, imagine that "a few Hogwarts staff members mistakenly take the Weasley twins George and Fred as the same person," meaning that in these staff members' belief worlds, the identity statement 'Fred = George' holds (Zhang 2016: 3). In such a scenario an identity statements like the following could felicitously be uttered, therefore: *Professor McGonagall thinks that Fred and George are the same person*. Hence, even though metaphysically Fred and George are not the same person, there exist certain worlds in which they are related via an equivalence relation. The truth of the relation 'Fred = George' is world-dependent, therefore.

4.2.2 Identity Statements in Counteridenticals

After having given arguments in favor of the hypothesis that in cross-world contexts, the semantics of identity relations are more complex than assumed from the given definition, since, for example, they might hold epistemically but not metaphysically, or the other way around, let us now consider how this insight helps us in understanding the relations which underlie the analysis of counteridenticals. To this end, let us start by considering

the case of pronoun resolution in the context of *If I were X*-conditionals.

4.2.2.1 Symmetric Identity Statements – Pronoun Resolution

Inspired by Frege’s discussion of attitude contexts, Lakoff (1996) has acknowledged the problem of non-rigidity of identity relations with respect to pronouns in the consequent of counteridenticals. The cases of counteridenticals and identity statements under attitude contexts seem related since in both scenarios, the truth of the identity statement is not evaluated against the facts of the actual world, but rather against those of designated belief worlds or counterfactual *p*-worlds, respectively.

For counteridenticals, Lakoff first and foremost questions the generative semantics’ principle to designate all first person pronouns as referring to the same person, i.e. the speaker, regardless of the different worlds their indexical content is intended to be resolved in by the speaker (cf. 91f). The identity relation which emerges from such a naive interpretation of first person pronouns can be schematized as ‘I/me/myself = the speaker.’ Put together with the Principle of Identity Substitution, this means that all occurrences of first person pronouns within a true statement should be substitutable by ‘the speaker’ and the result should remain to be true. The sentence ‘I washed myself,’ for example, turns into ‘the speaker washed the speaker’ under this principle, which seems to yield the same truth-conditions as the original sentence (cf. *ibid*). Despite the accuracy of this paraphrase, Lakoff argues that when considering counteridentical counterfactuals or dream reports, where there exist multiple references for first person pronouns, some of which “allude[] to an alternative world in which the speaker [...] is presented as experiencing something from someone else’s vantage point,” this pattern of inference is easily shown to be wrong (McCawley 1993: 417, cf. Lakoff 1996: 91f). In order to prove this failure of inference, let us have a look at the following example:

- (42) a. I/me/myself = the speaker.
b. If I were you, I would hate me.
c. If the speaker were you, the speaker would hate the speaker. ⚡

Lakoff shows that (42-c) cannot be a feasible paraphrase for (42-b) for two reasons: First of all, he argues that the Principle of Identity Substitution in combination with the identity statement in (42-a) would yield the same paraphrase for the sentence *If I were you, I would hate myself*, which, however, crucially differs in meaning to (42-b). Second, he adduces a syntactic argument: The substitution of all occurrences of first person pronouns by means of the same referent – the speaker – cannot be feasible for counteridenticals, since “a direct object is reflexive if and only if it is coreferential with its subject (except in certain well-known special constructions)” (92), meaning that whenever the direct object in the consequent clause is ‘me’ and not ‘myself,’ it should not be substitutable by the same

entity as the one the subject pronoun stands for in usual contexts, as otherwise a Principle B violation (cf. Chomsky 1981) would occur.¹⁴ In contexts where the counteridentical is about persons other than the speaker, e.g. “Mary said that if she were Susan, she’d hate herself,” such a substitution would lead to a Principle C violation instead (cf. *ibid*).

The fact that (42-b) nevertheless constitutes a valid counteridentical raises the question of how to deal with the first person pronoun reference in cross-world contexts, therefore. As a first intuition we might conclude that instead of there being one single identity relation which identifies the speaker as the referent of all first person pronouns in an utterance, there might exist a set of world-dependent identity relations that identifies each first person pronoun with its referent in a designated possible world. Since for each of these world-dependent identity relations between pronoun and referent, the Principle of Identity Substitution must hold in order for the resolution to work properly, we infer that the underlying identity relations must be symmetric as well as ‘epistemically’ true in all of the p -worlds, even though some of these relations might not hold metaphysically. Similar to Zhang’s (2016) Fred-and-George example, we conclude that we are dealing with world-dependent but otherwise trivial (i.e. conformable with the definition given in section 4.1) identity relations.

4.2.2.2 Asymmetric Identity Statements – Antecedent Relation

Now that we have explored the symmetric usages of identity statements, which we have argued to underlie pronoun resolution in the case of counteridenticals, let us turn our attention to their asymmetric counterparts, which we have predicted to provide the basis of the relation the antecedent clause of an *If I were X*-conditional sets up.

The idea of the asymmetric use of identity statement was also first discussed in the literature on attitude reports: it provides the second usage of identity statements in contexts where an identity relation does not hold in the actual world, but only epistemically (or counterfactually). In asymmetric identity statements, two entities which are not the same in the actual world are identified with each other under a mistaken identity context which has come to be since no sufficiently elaborate concept is available for at least one of the persons under discussion (cf. i.a. Cumming 2008; Percus and Sharvit 2014; see section 4.2.2.2). As a result, people in such mistaken identity contexts may believe that the expression $\llbracket a \text{ is } b \rrbracket$ is true while at the same time believing that the expression $\llbracket b \text{ is } a \rrbracket$ is false. In order to get an idea of how an identity statement can be used asymmetrically in an attitude context, consider the utterance (43-a) which occurs in reaction to the mistaken identity scenario in (43).

(43) MISTAKEN IDENTITY CONTEXT. Peter is throwing a party in honor of his cousin

¹⁴ Also see Lakoff (1996) for a similar discussion of pronoun reference in McCawley’s famous dream report example: *I dreamed that I was Brigitte Bardot and that I kissed me/myself*.

Dan who has just been awarded his PhD. All the guests know that, but they don't all know Dan (and some of them, like Kevin, don't even know the new PhD's name). When Becky arrives, Kevin, who is already completely toasted, walks up to her with a big smile. 'You must be proud to be a doctor now,' he says, 'is your wife coming too?' Seeing this, Jim says to Peter:

- a. 'Kevin thinks that Becky is Dan, (but he doesn't think that Dan is Becky).'
- (Zhang 2016: 4)

Intuitively, what the speaker wants to convey by uttering sentence (43-a) is not that Kevin thinks that Becky and Dan are in fact the same person; what he rather wants to convey is that Kevin is attributing Becky *de re* the concept which he has of Dan, i.e. being the new PhD, where *de re* means that the speaker is making a claim about a specific entity, a *res*, which is different from the speaker (here: Becky). Were the speaker to self-ascribe a property to himself, the expression would turn into a *de se*-statement instead of a *de re* one.

With regard to the topic of counteridenticals, the analysis of such asymmetric identity statements is deemed particularly revealing, since we recall from the previous chapters that the antecedent *If I were X* is not equivalent to *If X were me* but that both allow for different continuations to occur (cf. e.g. section 3.2.3). As a consequence, our intuition about the copular clause in counteridenticals is that we would not want 'to be' to be translated into the naive concept of identity as in 'Cicero = Tully,' even though such a concept is also able to correlate two actually distinct individuals, e.g. 'Fred = George,' which is needed for counteridenticals; instead, we want the copula to behave in a similar way as in mistaken identity contexts, where it transfers a substantial number of (contextually relevant) properties from one individual to the other. We have seen in the previous chapters that the number of properties which are being transferred differs widely between the 'advice' and the 'imagine' usages of identity statements.

Based on a similar empirical intuition for attitude contexts, Percus and Sharvit (2014) propose the following meaning for asymmetric copulas:

$$(44) \quad \llbracket \text{PRED} \rrbracket^{c,i} = \lambda k_{\langle s,e \rangle}. \lambda x_e. x = k(i)$$

The meaning of the copula is a two-place relation, which takes an individual concept k of type $\langle s, e \rangle$ as an input and identifies it with an individual x of type e . If such an individual concept, i.e. being the new PhD, is overtly available, the concept is simply predicated as a property of the subject referent (cf. (45)).

$$(45) \quad \underbrace{\text{Dan}}_{x_e} \text{ is } \underbrace{\text{the new PhD.}}_{k_{\langle s,e \rangle}}$$

$$\llbracket (45) \rrbracket^{c,i} = 1 \iff \llbracket \text{PRED} \rrbracket^{c,i} (\lambda s. \text{The new PhD student in } s)(\text{Dan}) = 1 \iff$$

$[\lambda k_{\langle s,e \rangle}. \lambda x_e. x = k(i)](\lambda s. \text{The new PhD student in } s)(\text{Dan}) = 1 \iff [\text{Dan} = \text{The new PhD student in } i] = 1$
 \Rightarrow *Sentence (45) is true at an index i iff Dan is the new PhD student in i .*

Sometimes, however, we do not want to identify a person with an overt individual concept but rather with another person, of whom we hold a certain concept. Such a scenario corresponds to the example of mistaken identity given above. In order to be able to account for such a situation, Percus and Sharvit suggest the following refinement of their semantics of asymmetric ‘to be:’

$$(46) \quad \llbracket \text{PRED } Z \rrbracket^{c,i} = \lambda x_e. x = f(\llbracket Z \rrbracket^{c,i})(i)$$

First, an individual (here: Dan) is coerced into an individual concept (here: being the new PhD) by means of the contextually salient type-shifter function $f: D_e \rightarrow D_{\langle s,e \rangle}$; $\llbracket Z \rrbracket^{c,i} \mapsto f(\llbracket Z \rrbracket^{c,i})(i)$. In a next step, a second individual (here: Becky) is identified with this individual concept by means of the asymmetric notion of the copula introduced in (44). In order for such an identification process between a person and another person’s context-dependent concept to work, Percus and Sharvit introduce the following constraint: $f(\llbracket Z \rrbracket^{c,i})(i)$ always denotes that individual in i who has certain properties that we presuppose $\llbracket Z \rrbracket^{c,i}$ to have uniquely. In other words: The type-shifter function f is always defined in a way such that the concept we derive from an individual at a certain world uniquely takes us back to that individual in that world, i.e. $f(\llbracket Z \rrbracket^{c,i})(i) = \llbracket Z \rrbracket^{c,i}$ (cf. Percus and Sharvit 2014).

Zhang (2016) criticizes Percus and Sharvit’s missing definition of the relation ‘=,’ which we know by now can be mistaken for equivalence (cf. sections 4.1 and 4.2.1). In order to circumvent this confusion, Zhang rewrites their proposal in terms of properties which can be checked to hold for an individual at a certain world. This redefinition reminds us of the predicational use of the copula ‘to be’ introduced in 4.1, which induced a set membership- or subset relation between individuals and properties instead of identifying two entities of the same kind in an equative manner.

$$(47) \quad \text{a. Redefinition of } \llbracket \text{PRED} \rrbracket^{c,i}$$

$$\llbracket \text{be}_{\text{asymmetric}} \rrbracket_{\langle \langle s,et \rangle, et \rangle}^w = \lambda P_{\langle s,et \rangle}. \lambda x_e. P(w)(x)$$

$$\text{b. Redefinition of } \llbracket \text{PRED } Z \rrbracket^{c,i}$$

$$\llbracket \text{be}_{\text{asymmetric}} \rrbracket_{\langle e,et \rangle}^w = \lambda y_e. \lambda x_e. P_{(w,y)}(w)(x),$$

where $P_{(w,y)}$ of type $\langle s, et \rangle$ represents the coercion of the individual y into some contextually salient set of properties in a world w .

(cf. Zhang 2016: 4f)

Adopting Zhang’s redefinition of Percus and Sharvit’s semantics leads to the following interpretation of (43-a):

$$\begin{aligned}
 (48) \quad & \llbracket \text{Kevin thinks that Becky is Dan} \rrbracket^{w_0} = 1 \iff \llbracket \text{be}_{\text{asymmetric}} \rrbracket_{\langle e, et \rangle}^{w_0}(\text{Dan})(\text{Becky}) \\
 & = 1 \iff \forall w' \in \text{Dox}_{w_0}(\text{Kevin})[P_{(w', \text{Dan})}(w')(\text{Becky})] \\
 & \Rightarrow \textit{Sentence (43-a) is true at } w_0 \textit{ if and only if every world } w' \textit{ in the set of Kevin's} \\
 & \textit{belief worlds is such that the contextually relevant properties of being Dan (here} \\
 & \textit{being a new PhD) hold for the res named Becky in } w'.
 \end{aligned}$$

(cf. *ibid.*)

From this calculation, we immediately see why the speaker who mistakenly believes that Becky is Dan does not necessarily have to believe that Becky is also Dan: Were we to switch the arguments in the calculation, we would need to coerce Becky into a set of contextually relevant properties making up a concept which the speaker then applies to Dan. Since in the context (43) no such concept is available for Becky, a similar calculation with the reverse order of arguments would not render a true statement.

From the above discussion, we see that in certain contexts, inherently individual-denoting entities can appear in the argument position of the predicational copula ‘to be,’ which induces a subset relation between the properties a person is coerced into (the concept of that person) and the properties of the person whom the concept is identified with. Given that this is clearly an asymmetric transfer of properties, it becomes evident that the order of the arguments fed into the copula function matters – a property which we have also observed for the antecedent clause of counteridenticals.

4.3 Concluding Remarks

All in all, we have seen that there exist two types of identity statements which are involved in the interpretation of counteridenticals: world-dependent, symmetric identity relations which underlie the resolution of the pronouns which occur in the conditional’s consequent clause, and asymmetric relations which provide the basis for the identification process induced by the structure’s antecedent clause. We predict that a valid theory of counteridenticals should incorporate both of these notions of identity.

Despite the detailed look which we have taken at identity statements in counteridenticals, there is one argument which we have not yet considered but which is crucial to both kinds of identity statements discussed: the cross-world rigidity of entities. We have seen from the asymmetric use of the copula that people possessing a mix of properties from two distinct individuals can arise in belief worlds and, by implication, also in counteridentical *p*-worlds: The Becky in Kevin’s belief worlds does not possess the same properties as the Becky who inhabits the actual world; instead, it seems that the new Becky is a mix of her properties from the actual world, subtracted by the contextually salient properties

regarding which the misconception has originated,¹⁵ and unified with the contextually salient properties of Dan (cf. (47)). If we adopt the notion that in order for two people to be identified as the same person at some index i , it is a necessary condition that the set of properties of these two individuals coincide at i , the Becky in the real world and the Becky in Kevin’s belief worlds should not be considered the same person. Nonetheless, let us agree with the literature on possible world semantics in that whenever we make a claim about another world, we implicitly also make a claim about the actual world as otherwise, our statements become meaningless, or, as Føllesdal (1986) puts it: “All our talk about [...] knowledge and belief, as well as about the other modalities, presupposes that we can keep our singular terms referring to the same objects. To the extent that we fail, these notions become incoherent” (111). As a result, we infer that the Becky in Kevin’s belief worlds should not only be closely related to the actual one but that we should also be able to understand that it is (some version of) *her* that we are talking about in a different world.

When looking back at Lakoff’s puzzle of first person pronoun reference in counteridenticals, we notice that a similar issue arises: Some of the first person pronouns seem to refer back to the actual individuals under discussion, making it easy to trace them, as well. Others, however, seem to refer to some “hybrid agent” defined by a mix of properties of the speaker and of the individual he is identified with by means of the identity statement in the antecedent clause (cf. Pelletier 2004: 151). In chapter 2 we have seen that the amount of properties the speaker discards in order to (partially) identify with the other entity can differ widely. Nevertheless, no matter how many properties the speaker counterfactually takes over from the object of the identity statement in the counteridentical’s antecedent, Føllesdal’s line of argumentation tells us that it only makes sense for the speaker to utter a first-person indexical in the consequent if he aims at making a claim about himself in some form, too.

All things considered, we thus conclude that there exists an overarching concept which underlies both kinds of identity statements relevant for the analysis of counteridenticals (those relevant for the pronoun resolution as well as the one identifying the set of relevant p -worlds), i.e. cross-world rigidity of entities.

¹⁵Note that such a subtraction of properties is crucial in order to prevent an individual to be defined via an inconsistent set of properties.

5 Theories of Counteridentical Conditionals

So far, we have sketched the prevalent theories of counterfactual meaning and we have argued that they are not trivially able to account for the different usages of counteridenticals. Furthermore, we have explored the range of meanings of identity statements in connection with their applications in counteridenticals. After having taken a close look at the two ingredients that make up the meaning of *counter-identicals*, we are now able to join our knowledge of the different fields and proceed to discuss different semantical analyses of *If I were X*-conditionals.

5.1 A Combinatorial Analysis?

From the structure of this thesis as well as a combinatorial semantics point of view, one would expect existing proposals of counteridentical meaning to straightforwardly combine the theories of their ingredient structures (i.e. counterfactuality and identity statements). Interestingly, however, this is not what we find; instead, the proposals to be discussed make only implicit use of the subcomponents' semantics, which is why the previous chapters can be understood to provide the theoretical background needed to understand the different approaches, as well as the formal toolkit which would be necessary to properly formalize them according to the principle of constitutionality for now (cf. chapter 1).

Several reasons why the theories may refrain from a straightforward combination of the semantic structures have already been mentioned in the previous chapters (in particular cf. section 3.2.3). Now we also have the means to buttress these claims in formal terms. Consider, for example, the following argument.

- (49) CONTEXT: My daughters work in the White House, which is why I do not get to see them often.
- a. If I_i were Michelle Obama _{j} , I_k 'd live in the White House and could see my _{i} daughters every day.

Let us assume that the meaning of the copula in the counteridentical's antecedent receives the asymmetric interpretation suggested by Percus and Sharvit and discussed in chapter 4.2.2.2. Furthermore, let us presuppose that the symmetric identity relation between the subject-*I* of the antecedent and the possessive-*my* in its consequent is realized through an unconventional binding relation which the syntax of the sentence takes care of: in doing so, we can either assume the subject-*I* from the antecedent to move directly to the topic position of the sentence, from which it can exert its binding power over the consequent pronouns directly; as another option, we could presuppose *my* to be an unbound pronoun, an assumption which we are acquainted with from the literature on donkey anaphora

(cf. e.g. Heim 1990; Kamp 1981; Chierchia 1995). With these assumptions in place, the following truth-conditions are obtained for sentence (49).

$$\begin{aligned} \llbracket (49) \rrbracket = 1 \text{ iff } \exists t' < t: \forall w' \in W [w' \in \text{SIM}_w (\text{HIST}_{w,t'}(\lambda w''. P_{(w'', \text{Michelle Obama})}(w''))(I))] \\ \longrightarrow w' \in \{w''' \mid [\lambda w''. \text{I live in the White House in } w''] (w''') = 1\}] \end{aligned}$$

\Rightarrow *Sentence (49) is true if and only if there is a time t' before the utterance time t such that all the worlds historically accessible at t' from the evaluation world w where the contextually relevant properties of being Michelle Obama (here, e.g., being the wife of the current president of the US with all of its implications) hold for the speaker and that are (overall) closest to w [according to SIM] are worlds where the thus defined counterpart of the speaker lives in the White House with the speaker's daughters.*

The problem which arises from these truth-conditions is of philosophical nature and arises in the same way for Iatridou's theory of counterfactuals: If the antecedent takes us to worlds in which the speaker has been (partly) identified with Michelle Obama, the p -worlds have been limited to a set of worlds in which presuppositions relevant for the interpretation of the sentence might not hold: for example, it is not at all clear that the actual speaker's daughters should exist in such worlds, since being the wife of the current president of the US might imply having children with the president instead. And even if their counterparts exist, meaning that the daughters mentioned are assumed to be of the speaker, it is not clear, given the assumptions about the speaker, that they would turn out the same way they have and, thus, end up working in the White House.

It becomes evident that the combination of the two subcomponents' theories does not yield the desired result. As a consequence, the proposals of counteridentical meaning which are going to be discussed in the next sections and which, to the best of my knowledge, exhaust the discussion of the meaning of *If I were X*-conditionals to this day – Lakoff (1996), Malamud (2006), and Kocurek (2016) – have opted to enhance the proposal by means of a further ingredient: the notion of cross-world rigidity. The importance of a strategy that keeps track of reference across worlds in the analysis of counteridenticals has been independently motivated in chapter 4.3. The best-known strategy to ensure such cross-world rigidity is Lewis' (1973) counterpart theory, which all of the three theories mentioned above constitute variations of. For this reason, they can be subsumed under the notion 'counterpart theories of counteridenticals,' even though they are located within different theoretical frameworks: Whereas Lakoff uses the framework of cognitive science/mental space theory, Malamud and Kocurek discuss the topic of counteridenticals from a semantics point of view.

5.2 Counterpart Theories of Counteridenticals

In the following, we will discuss Lewis' (1973) counterpart theory as well as the existing proposals of counteridentical meaning by Lakoff (1996), Malamud (2006), and

Kocurek (2016).

5.2.1 Preliminaries: Lewis (1973)

Lewis' counterpart theory tries to solve the puzzle of cross-world rigidity introduced above without assuming the existence of *transworld identity*, i.e. the possibility for one individual to inhabit more than one world, which he discards for philosophical reasons (cf. 39ff). Instead of presupposing identity of individuals across worlds, Lewis aims at satisfying Føllesdal's demand by making the following assumptions:

1. Each individual can inhabit exactly one world.
2. For each individual there may exist *counterparts* in other possible worlds, i.e. things inhabiting those worlds that resemble the individual under question "closely enough in important respects of intrinsic quality and extrinsic relations, and that resemble it no less closely than do other things existing there" (39).

From this proposal we conclude that whatever an individual is or does at a world different to the actual one, he neither is nor does himself but rather "vicariously through his counterpart" (ibid.). With regard to the topic of counteridenticals, such a concept seems appealing since it implies that the referent of the *I* in the actual world is necessarily different from the one(s) of which the consequent proposition holds, which seems especially important in the case of the 'hybrid agent' introduced before.

Based on Lewis' definition of counterparts, individuals are usually assumed to possess either one or zero counterpart at each possible world, since there either exists an entity which is sufficiently close to it there or not (cf. ibid); Nevertheless, since the counterpart relation is defined in terms of similarity, which we have made out to be a notion that is not well-defined in section 3.3, there are certain contexts in which several counterparts may exist for the same entity in the actual world. As a logical conclusion, there also exist more than one counterpart relation in such cases. To illustrate such a scenario, let us briefly think back to our toy examples for proving the vagueness of the similarity relation from before, restated here as (50-a) and (50-b)

- (50)
- a. If Caesar had been in command in Korea, he would have used the atom bomb.
 - b. If Caesar had been in command in Korea, he would have used catapults.

Even though the antecedent takes us to the same set of worlds, i.e. those at which Caesar was in command in Korea, in the first situation, Caesar's counterpart at those worlds is derived via a similarity relation based on the emperor's character whereas in the second case, it is obtained via a counterpart relation that stresses the historical facts that held at the emperor's lifetime.

On the note of the counterpart relation's flexibility, it is Lewis' discussion of the following pair of sentences that is particularly interesting with regard to the analysis of *If I were X*-conditionals.

- (51) a. If New York City were in Georgia [emphasis added], New York City would be in the South.
 b. If Georgia included New York City [emphasis added], Georgia would not be entirely in the South.

(cf. Goodmann, ctd. in Lewis 1973: 43)

Here, the same two entities are identified with each other by means of the same relation in both of the counterfactuals' antecedents – even though in opposite order –, the two consequent clauses are contradictory under the presupposition of the antecedent clause's truth, and still, both sentences can be true at the same time – a pattern which we also find with regard to counteridenticals:

- (52) 'Imagine'
 a. If I were Michelle Obama, I'd live in the White House.
 b. If Michelle Obama were me, she'd live on Oxford Street.

- (53) 'Advice'
 a. If I were you, I'd go skydiving tomorrow.
 b. If you were me, you wouldn't go skydiving tomorrow.

Lewis (1973) justifies the simultaneous felicity of such pairs of sentences by means of a variance "in stringency of resemblance" which the relevant counterpart relations demand (42f). For each sentence in such a pair, one of the two counterpart relations ensuring the cross-world rigidity of the entities under discussion is always stricter than the other with respect to the facts of the world. In the examples given above, I have underlined the entity whose counterpart relation is (contextually relevantly) stricter (cf. Lewis 1982: 43).

Even though the highlighting may suggest that we can derive a pattern with respect to the different usages of counteridenticals in this matter, i.e. in 'imagine' counterfactuals it is always the object counterpart relation which is stricter whereas in 'advice' contexts the stricter counterpart relation holds of the subject, a change in focus brings about a shift of the strictness of the counterpart relation for the 'advice' reading, for example:

- (54) CONTEXT: You are afraid of heights and have been invited to go skydiving tomorrow. I am not afraid of heights at all and love doing adventurous things. You have asked me what I think you should do.]

- a. If $\overset{F}{I}$ were you, I'd go skydiving tomorrow. (I'd love that!)

- b. If I were \overline{you} ^F, I wouldn't go skydiving tomorrow. (You'd hate it!)

In these examples, we see that it is always the focused entity whose counterpart is derived via a stricter counterpart relation, which, in (99-a) takes seriously the speaker's love of adventure and, in (99-b) takes seriously the addressee's fear of heights.

All things considered, Lewis' counterpart theory is appealing conceptually with respect to cross-world rigidity of entities, and has the potential to capture the meaning of *If I were X*-conditionals. In fact, Lewis himself acknowledges the usefulness of his theory with respect to the analysis of counteridenticals:

For a familiar illustration of the need for counterpart relations stressing different respects of comparison, take '*If I were you ...*'. The antecedent-worlds are worlds where you and I are vicariously identical; that is, we share a common counterpart. But we want him to be in your predicament with my ideas, not the other way around. He should be your counterpart under a counterpart relation that stresses similarity of predicament; mine under a different counterpart relation that stresses similarity of ideas.

(Lewis 1973: 43)

Despite the theory's usefulness with respect to the analysis of counteridenticals, which all of the existing proposals of counteridenticals to be discussed in the next chapter also exploit, it should be acknowledged that the basic assumptions it is built on are not uncontroversial in the literature.

Pollock, for example, criticizes the fact that Lewis' definition of counterparts relies on the assumption that the counterpart a' of a in a possible world w' "bears at least a minimal similarity to a and is more similar to a than is any other object" in w' , even though sometimes we want to hypothesize about situations in which a' is crucially *not* too similar to a but where there, instead, exists another object b' which is very similar to a (Pollock 1976: 110; cf. Lewis 1973: 39). In such situations, Pollock argues, the counterpart theory fails as it excludes worlds in which such relations hold from the set of p -worlds. He gives the following example:

- (55) 'If I had been born in the place of Richard Nixon, with all of his genes, etc., and raised as he was raised, and he in turn had been born and raised in my place, then I would have been a president threatened with impeachment and he would have been an interested bystander.'

(Pollock 1976: 110)

Here, the counterfactual p -worlds are worlds in which the speaker's counterpart is more similar to Richard Nixon than is his actual counterpart in those worlds; Richard Nixon's counterpart, on the other hand, is more similar to the speaker than is the speaker's

counterpart. As a result, Pollock concludes that Lewis' counterpart relation needs to be refined in a way that allows for such sentences to be acceptable.

Nevertheless, it is self-evident that the loosening of Lewis' definition of counterparts would demand the identification of a "new criterion of counterparthood" (ibid.: 111) or a new theory of cross-world rigidity altogether. To this day, there does not exist a suitable alternative theory, which would solve this problem, since the possibility to trace an individual across worlds, which we have seen is crucial for the analysis of counteridenticals, is dependent on the assumption of some kind of essence of entities, which is generally a topic highly debated in the literature (cf. e.g. Chrisholm 1967).¹⁶ This lack of a better solution motivates me to acknowledge Pollock's criticism but at the same time ignore it for the remainder of this thesis.

5.2.2 Lakoff (1996)

Lakoff's proposal for the meaning of counteridenticals is based on his assumption that each individual can be conceptualized as the combination of two separable parts: the individual's *Subject* and his *Self*. In this context, the Subject of a person is assumed to be the individual's "locus of subjunctive experience: consciousness, perception, judgment, will, and capacity to feel," whereas the entity's Self can – simplistically – be understood as the individual's metaphysical representation, i.e. his body, and the external conditions connected to it (for example, the individual's past and social role) (93f). With regard to *If I were you*-counteridenticals, where two individuals are identified with each other, this means that we have to take into account four distinct entities: The *Subject-of-I*, *Subject-of-You*, *Self-of-I*, and *Self-of-You* (cf. ibid).

Building on Lewis' counterpart theory, Lakoff first suggests that the meaning of *If I were you*-conditionals comes about via a counterpart relation which creates a hybrid agent in the counterfactual *p*-worlds which combines the *Subject-of-I* from the actual world with the *Self-of-You* from the actual world via identifying the two Subjects at hand with each other across worlds (cf. Lakoff 1996: 93f). Such an interpretation of the counteridentical antecedent clause directly corresponds with Lewis' proposal that in the *p*-worlds of *If I were you*-conditionals, the speaker and the addressee share a common counterpart, which is in the addressee's predicament with the speaker's ideas (cf. Lewis 1973: 43 qtd. in chapter 5.2.1).

Lakoff's idea of counterpart semantics was mainly pursued further in the field of cognitive semantics/mental space theory (cf. Fauconnier 1985), which is the framework Lakoff eventually also adopted instead of formal logic (cf. Lakoff 1996: 93). The general

¹⁶ According to Kripke (1980), for example, even though we might be cancel any of an individual's properties (e.g. Peter is tall. *If Peter weren't tall, . . .*), at least the name of an individual should rigidly refer to the same individual across worlds. Nevertheless, the following example shows that this assumption, too, is not undisputable: *If Peter weren't \overline{Peter} , he would go out with us tonight.*

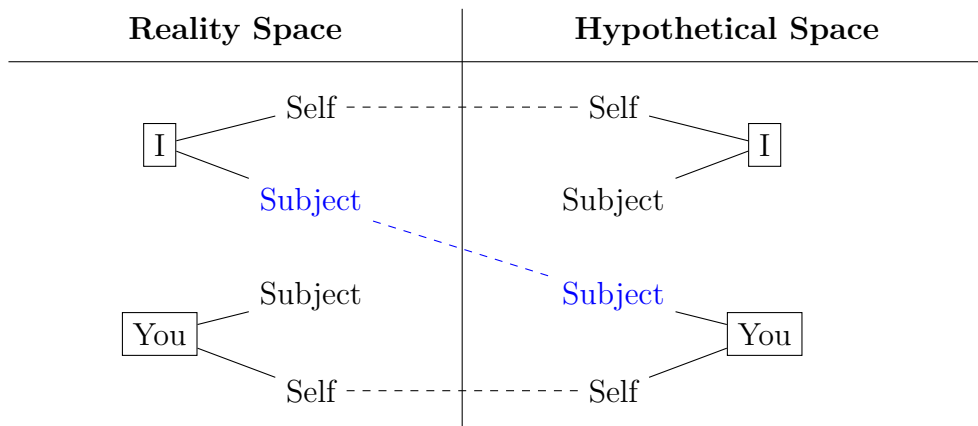


Figure 4: Lakoff's model of *If I were you*-conditionals (cf. 94)

idea of his approach remained the same, but the terminology and the related concepts changed, meaning that instead of his proposal's building on technically challenging counterpart relations, in his theory, cross-world equivalents are defined in terms of conceptual "connectors," and instead of assuming the existence of possible worlds, he argues in favor of the existence of hypothetical mental spaces (cf. Lakoff 1996: 93). As a consequence from this change in framework, Lakoff proposes the following schema as underlying *If I were you*-conditionals:

Here, the antecedent clause *If I were you* is responsible for doing two things: First of all, as it includes a contrary-to-fact proposition, it sets up a hypothetical space that exists parallel to the reality space. Secondly, it induces the creation of a connector (represented in the model by dotted lines) which links the speaker and the addressee across the two mental spaces in such a manner that it identifies the speaker's Subject in the reality space and the addressee's Subject in the hypothetical space with each other. This identification leads to the genesis of a counterpart of the addressee in the hypothetical world which retains the addressee's original predicament, while at the same time incorporating the speaker's locus of subjunctive experience instead of the addressee's (cf. *ibid*). The truth of the conditional's consequent clause is then evaluated with respect to this hypothetical space and the individuals inhabiting it.

What we have just seen schematically can be formalized as follows:

- (56) Lakoff's proposal for a sentence of the form "If NP₁ were NP₂, NP₃ would VP" (Constraint: NP₃ is an anaphora and NP₁ is its antecedent)
- a. there are two mental spaces, the Reality Space, R , and a Hypothetical Space, H , dependent on R ;
 - b. the referents of NP₁ and NP₂ are in R , and the referent of NP₃ is in H ;
 - c. each referent of an NP is conceptualized as having a Subject and a Self;
 - d. NP₃'s Subject is the counterpart of NP₁'s Subject. NP₃'s Self is the counter-

- part of NP₂'s Self;
- e. VP predicates the Subject properties of NP₃ that result from NP₂'s Self being paired with NP₁'s Subject;
 - f. NP₂ VP is false in *R*; NP₃ VP is true in *H*; and
 - g. the antecedent-anaphora relationship indicates not full person identity, but rather Subject identity between NP₃ and NP₁.

(Lakoff 1996: 95)

Lakoff's theory presents an easily comprehensible implementation of Lewis' counterpart theory, which furthermore makes the identity statements assumed for the analysis replicable. For this reason, it constitutes a suitable starting point for our analysis. Nonetheless, Lakoff assumes two constraints regarding *If I were you*-conditionals which seem dubitable in light of the results from the typological study laid out in chapter 2. These constraints are the following.

1. NP₃ cannot refer anaphorically to NP₂.
2. VP must predicate Subject properties, not Self properties.

When recalling the typological investigation in section 2, we realize that these restrictions directly coincide with those we have made out to hold for the 'advice' use of counteridenticals, which is why we predict Lakoff's theory to be able to capture their meaning correctly. Let us briefly and informally test this hypothesis with regard to our usual toy sentence for the 'advice' use of counteridenticals:

(57) If I_i were you _{j} , I_i 'd buy the blue dress.

Here, the speaker is projecting his subjective properties into the predicament of the addressee, i.e. being in the situation of deciding which dress to buy, which leads to a Subject identity between the speaker in the reality space and the addressee in the hypothetical space. Hence, the proposed antecedent-anaphora relation holds (cf. (56-g)). What is more, the VP in the *q*-proposition talks about NP₃'s judgment, which, by assumption, are the speaker's judgments. Since judgment is part of the set of Subject properties, we see that the second restriction Lakoff proposes holds for this sentence, as well.

Let us turn our attention to the second case of counteridenticals, i.e. those which are applied in their 'imagine' interpretation, and let us investigate the validity of the restrictions 1. and 2. in this context. Lakoff argues that only the subject of the antecedent clause, NP₁, can function as the antecedent of the anaphoric expression NP₃. In order to buttress his claim, Lakoff adduces the following pair of sentences:

(58) a. If I_i were you _{j} , I_i 'd get upstairs this second! [indices added]

- b. *If I_i were you_j, you_j'd get upstairs this second! [indices added]
 (Lakoff 1996: 95)

Without doubt, sentence (58-b) sounds odd and its contrast with (58-a) proves that this oddness is grounded in the use of the second person pronoun in the counterfactual's consequent clause. Nonetheless, I do not conclude with Lakoff on the basis of the minimal pair in (58) that this oddness is due to the antecedent-anaphora relationship induced by the pronoun; rather I suggest that it comes about by the use of the second person pronoun itself. The following variant of sentence (58) serves to illustrate my hypothesis:

- (59) CONTEXT: We broke my mom's favorite vase while playing and know that she is going to be very mad about this. As we stand in the hallway, trying to figure out what to do, we hear her car approach. You ask me what I am going to do now and I respond:
- a. *If I_i were ^Fyou_j, I_j'd get upstairs this second!* You always run away from responsibility. I, however, will just stay down here and explain to her what happened. I'm sure she'll understand.

Lakoff describes the antecedent-anaphora relation as indicating 'not full person identity, but rather Subject identity between NP₃ and NP₁' (cf. (56-g)). From the above example, however, we see that Subject identity can also hold between NP₃ and NP₂, which leads us to the conclusion – contra Lakoff – that both NP₁ and NP₂ can function as NP₃'s antecedent.

A further restriction Lakoff poses on counteridenticals is that only Subject properties can be predicated of the hybrid agent in the counterfactual world and not Self properties. This insight, he argues, can be made explicit by considering examples like the following:

- (60) a. *If I were you, I'd be short and named George.
 b. *If I were you, I'd hate brussels sprouts. (construed as a physical property)
 (Lakoff 1996: 96f)

Even though Lakoff withholds from making this restriction an ultimate rule (he states that in "far-fetched contexts" involving "judgment or emotionality" sentences like (60-b) might be acceptable (98)), we have seen from the previous sections that, in fact, *any* restriction of this kind would be too strict, since – together with the constraint that only NP₁ can function as the antecedent of NP₃ – it would rule out all 'imagine' readings by default. However, we have seen that there exist numerous 'normal' contexts in which the 'imagine' use of counteridenticals can be applied felicitously (cf. the examples in section 2).

On a different note, I want to point out a further problem that arises with regard to

Lakoff's theory of mental spaces for the analysis of counteridenticals: When we look at the design of Lakoff's hypothetical space, we notice that the only individual inhabiting it is the counterpart of the addressee – made up of the addressee's actual predicaments and the speaker's actual subjunctive properties. One might argue that the notion *I* also exists in this space; nevertheless, it is unclear to me if the *I* in the hypothetical space should be counted as the speaker's counterpart, or even an individual, since it does not possess a Subject. The question which immediately presents itself from this observation is how to deal with multiple referents for the first person pronouns in the counterfactual's consequent as discussed in sections 2.2.2 and 3.2.3.

There is more that could be said about Lakoff's theory. For example, it further seems questionable whether his restriction (56-f) holds, since one might suggest something to the object of an 'advice' counteridentical which the object itself might have done, as well, ruling out the falsehood of 'NP₂ VP' assumed by the constraints. Nevertheless, due to the restraints of this paper, I will leave the discussion at this point and conclude that Lakoff's theory does not uphold as a theory of counteridenticals in light of our findings from section 2.

5.2.3 Malamud (2006)

We have seen that the main fault of Lakoff's theory is its inability to account for the 'imagine' use of counteridenticals. In the following, we will take a closer look at Malamud's proposal for *If I were X*-conditionals, which will turn out to be a more flexible version of Lakoff (1996), translated back into the framework of semantics. Before examining her analysis, however, it should be mentioned that Malamud (2006) was not intended as a theory of counteridentical meaning in the first place, but that the analysis rather emerged as a side product from Malamud's investigation of the semantics of arbitrary pronouns, like English *you*. With this in mind, let us now turn our attention to her analysis.

Similar to Lakoff, Malamud, too, conceptualizes individuals as the combination of two separable parts, which she calls *Self* and *Persona* (cf. 92); and, similar to Lakoff, Malamud also bases this conceptualization on Lewis' counterpart theory (ibid.). In contrast to him, however, she makes use of Lewis' theory within the framework of formal semantics, meaning that she tries to incorporate the notion of counterparts into the semantical representation of the counterfactual antecedent instead of assuming external, conceptual 'connectors' across worlds. What is more, despite of the similar terminology both theories apply in order to refer to the two entities which make up an individual, the terms connote vitally different meanings in the proposals. For this reason, let us be explicit about how Malamud uses the different notions before we take a look at the sample sentences:

(61) Definitions

- a. *Self of x*: the bundle (conjunction) of properties *P*, such that $\Box P(x)$ - i.e.,

in every world w accessible from the actual world, counterparts of x have the properties P .

- b. *Persona of x from y* : a subset of properties Q of y that in some accessible worlds is true of a counterpart of x . *Persona* thus is a two-place predicate, relating two individuals in a world: the actor x and the role-provider y .

(Malamud 2006: 92)

Whereas Lakoff conceptualizes individuals as the combination of their internal (*Subject*) and external (*Self*) properties, we see that Malamud's splits them according to a different criterion: In her theory, the notion *Self of x* constitutes an individual's set of *necessary* properties, i.e. properties which hold for every counterpart of that individual across all worlds. The other entity an individual can – simplistically – be broken down into is his *Persona*. The term 'simplistically' here points to the fact that *Persona* is a relation which actually takes two individuals as inputs and relates them in the fashion described in (61-b). Nevertheless, the term's definition suggests that a person x can still be split into his *Self-of- x* and *Persona-from- $x(y)$* , since the latter can be comprehended as a function which first coerces x into a certain set of (contextually relevant) properties, Q , and then takes a second individual, y , as its input and takes us to those worlds w_i where Q is a subset of the properties defining y_i ($:=$ the counterpart of y in w_i). Hence, an individual in Malamud's approach is composed of his 'essence' and his contextually relevant properties; from this definition we infer that only the union of all possible *Personas*, i.e. sets of contextually relevant properties ("role[s]") (ibid.: (92)) that an individual can provide together with his essential properties is able to make up the entire set of properties that define the person in the actual world (cf. ibid.).

Malamud's conceptualizing individuals in a different way than Lakoff leads to her being able to make the following refinement to their shared concept of split individuals: "the particular subset Q picked out by the persona relation varies widely depending on the sentence" (92). By means of this small adjustment, Malamud argues, her approach is able to account for both of the following sentences, which can easily be made out to correspond to the 'advice' and 'imagine' use of counteridenticals, respectively:

(62) Variation in personas

- a. If I were President Bush, I would not have started the war in Iraq.
- b. If I were President Bush, I would be a raving war-monger.

(Malamud 2006: 92)

From the discussion of Lakoff's proposal, we recall that his theory was unable to capture the correct meaning of (62-b), since it always identifies the individuals in the counteridentical's antecedent clause in such a way that their common counterpart possesses the Subject of NP_1 and the predicaments of NP_2 . Clearly, however, in (62-b), the property of being a

raving war-monger is derived from the Subject properties of NP₂.

With this in mind, let us now look at a sample derivation of a sentence’s truth-conditions under Malamud’s (2006) proposal, which, for reasons of simplicity, omits “such complications as modal base and ordering source” (93):

$$(63) \quad \llbracket \text{If I were Mary, I wouldn't be dating that horrid guy} \rrbracket^{c,w} = \\ \lambda w_0. \forall w [\text{persona}(\text{Mary}, \text{speaker}(c), w)] [\neg \text{date}(\text{speaker}(c), \text{the-horrid-guy}, w)]$$

The truth-conditions of this counterfactual can be paraphrased as follows: “For all worlds w accessible from w_0 , the speaker, inhabiting Mary’s persona, does not date the contextually salient horrid guy.”

There are several things worth discussing about Malamud’s approach in general as well as about this derivation in particular: First of all note that the conditional in (63) is ambiguous between an ‘advice’ reading and an ‘imagine’ reading, depending on the context.

- (64) a. ‘Advice’:
 Mary is so great and Tom is such a jerk. I don’t know why she doesn’t leave him. *If I were Mary, I wouldn’t be dating that horrid guy.*
- b. ‘Imagine’:
 I’m so angry at Tom, he’s always bringing me down. I can’t leave him, though, I’m too afraid of being alone. I wish I was as independent as my sister, Mary. *If I were Mary, I wouldn’t [still] be dating that horrid guy.*

Due to the ambiguity of the sample sentence, it provides a suitable opportunity to test Malamud’s claim of her theory’s being able to achieve both readings via a variation of persona-relations. Self-evidently, Malamud’s hypothesis holds if both of the correct interpretations of (63) can be obtained via different sets of properties Q which hold of Mary in the actual world and which the speaker counterfactually imagines to take over. The correct meaning for (64-a) is achieved if Q is taken to only include properties related to Mary’s predicament and none of her contextually relevant, subjunctive properties, whereas the intended interpretation of (64-b) comes about less restrictedly, i.e. if Q either includes Mary’s internal properties in addition to the properties related to her predicament or not. This flexibility is due to the fact that it does not matter whose mindset the hybrid agent in the counterfactual world possesses as long as his external situation is that of the speaker’s sister. Since, in contrast to Lakoff (1996), these different sets of properties can easily be accounted for by Malamud’s theory, at a first glance, it seems promising with regard to the research question of this thesis.

Nevertheless, a closer look at the approach reveals that like Lakoff’s theory it, too, faces several problems. The first of Malamud’s assumptions which seems questionable

becomes evident when looking back at the formal account of the semantics of (63). Here, we see that there are two individuals in debate (Mary and the referent of the pronoun *I*), but only one of them (Mary) is split into *Self* and *Persona*. Even though one might argue that in this specific case it does not matter that the referent of the pronoun *I* is not split as such, as the asymmetry in the consequent seems self-evident under normal prosody, Malamud argues in her proposal that there generally exist restrictions with regard to which people are specified for *Persona* and, thus, function as a role provider, and which are not: The pronoun *I*, for example, is an exception to this rule according to Malamud, meaning that its semantics always amount to the following:

$$(65) \quad \llbracket I \rrbracket^{c,w} = \text{speaker}(c) \quad (\text{Malamud 2006: 93})$$

This premise seems dubitable for several reasons: First of all, the interpretation of all pronouns in counteridenticals as referring to the speaker leads us back to the problem of the predicted synonymy of the sentences *If I were you, I would hate me./If I were you, I would hate myself.* discussed in length in 4.2.2.1. What is more, the theory's ban on the split of the speaker into *Self* and *Persona* means that the referent of the pronoun *I* in the actual world cannot function as a role provider, which we know is not accurate since sentences like (52-b) and (53-a) are feasible counteridentical statements. Lastly, there is no obvious rationale as to why some individuals should be separable into two entities whereas others are not, or – more specifically – as to why individuals per se are separable into two entities, but as referents of first person pronouns, they are not.

This rationale only becomes evident when remembering that Malamud's proposal emerged as a side product of her analysis of impersonal pronoun uses: Since Malamud adduces the divisibility of pronouns like *you* as her main argument as to why they can be used both impersonally and personally, whereby its specification for *persona* ensures the impersonal use, she has to assume the inseparability of the referent of the pronoun *I*, which she argues cannot be used in the impersonal way (see Zobel (2014) for a refutation of this hypothesis). Consequently, she has to rule out the separability of the pronoun *I* to stay consistent in her analysis. For a more in-depth analysis in this regard, the reader is advised to consult Malamud (2006).

In addition to the problem discussed above, like Lakoff (1996), Malamud's proposal is also unable to account for the existence of multiple referents for the first person pronouns which can occur in a counteridentical's consequent clause: As per Malamud, the antecedent clause induces a counterpart relation which equips the subject of the clause, here *I* – i.e. the speaker –, with a certain set of contextually relevant properties, *Q*, of the clause's object. By this means, Malamud is able to use the notion 'speaker(*c*)' in the analysis of her consequent clause to refer to the hybrid agent, since the aforementioned relation has redefined the notion for the counterfactual world, i.e. it has set up an appropriate

counterpart. Since, by this means, any mentioning of ‘the speaker(*c*)’ in the consequent of the conditional necessarily picks out the thus defined counterpart of the utterance’s speaker, Malamud cannot account for sentences in which first person pronouns have multiple referents as in sentences like (37), restated here as a reminder as (66).

(66) If I were you, I’d feed my cat every day.

$I_{as\ A} \quad I_S$

Lastly, Malamud’s theory relies on the existence of an individual’s essential properties, since she takes the Self of an individual to consist of those properties which hold of the relevant individuals across worlds ($\Box P(x)$). Nevertheless, the discussion of Lewis’ counterpart theory has proven that the existence of an individual’s essence is a highly debated topic in the literature, since there does not seem to exist any property which could not be canceled of an individual by means of a counterfactual antecedent (cf. chapter 5.2.1). It has been pointed out to me by von Fintel (p.c.) that a possible remedy to this challenge of Malamud’s theory could lie in the assumption that the essence of a person cannot be manifested in language.

All in all, we conclude that we have not yet found a suitable theory of counteridenticals in Malamud’s approach.

5.2.4 Kocurek (2016)

The most recent as well as the most elaborate counterpart theory of *If I were X*-conditionals to this day has been put forward by Kocurek (2016). It sets out to remedy the problems the other counterpart theories have encountered, i.e. accounting for both the ‘advice’ as well as the ‘imagine’ use of counteridenticals, incorporating the possibility to derive counterparts for each individual in the actual world, and accounting for multiple referents of the first person pronouns in the counteridenticals’ consequent clause, while still keeping Lewis’ basic assumptions in place. The manner in which Kocurek aims at achieving this remedy is via taking seriously the idea that each entity may have several distinct counterparts depending on the context (cf. chapter 5.2.1). Hence, he believes that not only every entity in the actual world, but every mentioning of such an entity should receive an individual counterpart index, whose value as well as content is to be resolved by context (cf. Kocurek 2016: 22). By assigning each occurrence of an entity/pronoun its own counterpart index and then quantifying over these counterparts, Kocurek lays the foundation for obtaining the correct meaning of complex reference structures of pronouns such as (66), as well as for providing a solution to the other problems stated above.

Kocurek bases this hypothesis on Lewis’ proposal that counterparthood is based on the notion of similarity, meaning that the consideration of the same individual under different notions of similarity will generate different counterparts of that person, which

may even exist within the same world (cf. chapter 5.2.1; Kocurek 2016: 22). It is further Lewis from whom Kocurek extrapolates the idea of how to find out which counterpart indices assigned to the different terms should match: According to Lewis, it is the “sense of a term” (Kocurek 2016: 23) which determines the counterpart relation that should be attached to it. Since different terms may coincide in the sense in which they are used, these terms should be assigned the same index to explicate that they have been derived under the same notion of similarity (cf. Lewis 1971: 209). Lewis further argues that sometimes the sense of a term is already explicitly determined by means of a designated clause which modifies the term (cf. *ibid.*: 210). Kocurek takes up the idea of making explicit the counterpart relations which should be attached to a term and suggests that they can be inquired via testing which modifier clause of the type ‘as ___’, ‘being ___’, etc. captures the sense of a term in a given context most appropriately (cf. Kocurek 2016: 24).

On the basis of these assumptions, Kocurek then developed the following algorithm for analyzing the semantics of counteridenticals:

1. Regiment the English sentence without any counterpart indices.
2. Go term by term and determine what phrase of the form ‘as ___’, ‘with ___’, or ‘being ___’ (or others like ‘*qua* ___’, etc.) would be appropriate to modify that term with in that context.
3. Go back and assign counterpart indices to terms and counterpart relations to counterpart indices that would make sense of those phrases. Hereby, assign a separate index to each distinct way of filling in the blanks in 2.

(cf. Kocurek 2016: 24)

In addition to the truth of the above algorithm, Kocurek tentatively assumes that the copular clause in the antecedent of counteridenticals is equative, i.e. symmetric. In his reasoning, which is contra that laid out in chapter 4.2.2.2 of this thesis, he adduces arguments like the infeasibility of the coordination of the ‘predicate’ in the clause *If I were X*: Usually, it is allowed to coordinate different predicates with each other, meaning that instead of assigning these properties to a given individual one by one, they may be clustered into a set of properties which is then related to the entity by means of a single predicative copula (e.g. {*Cicero is smart. Cicero is kind. Cicero is brave.*} $\xrightarrow{coor.}$ *Cicero is smart, kind, and brave.*) Interestingly, such a clustering is infeasible if one of the predicates to be coordinated is derived from a counteridentical statement (i.e. {*Cicero is smart. Cicero is well-read. Cicero is Tully.*} $\xrightarrow{coor.}$ **Cicero is smart, well-read, and Tully.*) (cf. Kocurek 2016: 7). Even though Kocurek argues in favor of an equative interpretation of the copula clause, he also lists several good arguments in favor of the predicative and against the equative analysis. This motivates him to admit the defeasibility of his assumption – a defeasibility which we will further try to buttress later on, in the spirit of section 4.2.2.2.

First, however, let us take a look at how Kocurek’s proposal works in action by considering the following ‘advice’ counterfactual:

(67) If I were you, I would bet on that horse.

To this end, let us follow the algorithm introduced above step by step, taking into account the assumption of the copula clause’s equative interpretation.

1. If I were you, I would bet on that horse.
2. If I (with my beliefs) were you (as a bettor), I (with my beliefs) would bet on that horse (being the fastest horse here).
3. $(I^1 = you^2) > Bet(I^1, that\ horse^3)$.
 \Rightarrow *Sentence (67) is true if all the closest worlds where a C_1 -counterpart of the speaker is a C_2 -counterpart of the listener, that counterpart will bet on every C_3 -counterpart of that horse; where C_1 might be a belief-counterpart relation, C_2 the predicament-counterpart relation, and C_3 the body-counterpart relation.*

(cf. Kocurek 2006: 22ff)

When recalling chapter 2, where we made the different readings of an ambiguous sentence explicit via specifying pronouns by means of phrases like ‘as S’, ‘as A’, etc., Kocurek’s analysis seems very promising – at least at a first glance.

In order to see if Kocurek’s analysis is able to remedy the problems the other counterpart theories have run into, let us consider them case by case. To this end let us first observe without a detailed line of reasoning that Kocurek does not exclude any individuals as possible role providers in the sense of Malamud, since he conceptualizes every individual in a context-dependent sense without any exceptions.

Next, let us see how he deals with the problem of multiple referents for the first person pronouns in the counterfactual’s consequent clause. For this purpose, consider the following minimal pair:

- (68) a. If I were you, I would vote for me.
 b. If I were you, I would vote for myself.

Via assigning (mis-)matching counterpart indices to the two occurrences of first person pronouns in the conditionals’ consequents, Kocurek (2016) argues that we can explicate the sense in which *I* and *me* in (68-a) refer to different people, whereas the pronouns *I* and *myself* in (68-b) have the same referent (cf. 25).

(69) Kocurek’s analysis of (68)

- a. $(I^1 = you^2) > VoteFor(I^1, me^3)$.
If I (with my beliefs and preferences) were you (as a voter), I (with my beliefs and preferences) would vote for me (as a candidate).

- b. $(I^1 = you^2) > VoteFor(I^1, myself^1)$.
If I (with my beliefs and preferences) were you (as a voter), I (with my beliefs and preferences) would vote for myself (with my beliefs and preferences).
 (ibid.)

Due to Kocurek’s theory allowing for the same individual to be interpreted in many different senses, in contrast to Lakoff and Malamud, it is able to predict a difference in meaning between the two sentences. This observation might lead us to conclude that his proposal is superior to the previous theories in this regard, too.

Lastly, let us take a look at the distinction between the ‘advice’ and the ‘imagine’ reading of counteridenticals. In doing so, let us exploit the following fact for *If I were X*-conditionals in their ‘imagine’ use: If a claim holds of the speaker under full identity with a person *X*, it necessarily also holds of the speaker under partial, contextually relevant identification with *X*. For example: The consequent clause ‘I’d live in the White House.’ is true in those possible *p*-worlds in which I am actually imagining to be Michelle Obama herself (strict interpretation of ‘If I were Michelle Obama’); by necessary implication, it is therefore also true in worlds in which I am only taking on Michelle Obama’s current housing situation (loose interpretation of ‘If I were Michelle Obama’). This inference is adequate since Michelle Obama’s housing situation constitutes a subset of the set of properties defining her, meaning that the *p*-worlds in which complete identification is assumed form a subset of those in which the speaker is only assumed to take on a subset of Michelle Obama’s properties. As a result, each claim which is true for the more restricted interpretation of *p*-worlds also holds of the less restricted one.

Hence, the ‘imagine’ counterfactual ‘If I were Michelle Obama, I’d live in the White House.’ can be paraphrased in the following way for the purpose of this analysis:

(70) ‘If I were identical to Michelle Obama, I’d live in the White House.’

For this sentence, Kocurek’s (2016) analysis predicts the following analysis (cf. 29):

(71) $(I^1 = Michelle\ Obama^1) > Live(I^1, White\ House^2)$.

From these semantics we see that Kocurek differentiates between ‘advice’ and ‘imagine’ counteridenticals by means of mismatching and matching counterpart indices in the antecedent clause, respectively. Note that the assumption of full identity was only made for reasons of keeping the analysis simple, and since Kocurek specifically assigns semantics for ‘identical to’-conditionals. Self-evidently, any notion of predicament-counterpart relation could be substituted for the index 1.

All in all, we conclude that Kocurek’s theory seems appealing not only conceptually as it conforms with our intuition of being able to disambiguate sentence meanings via

designated modifying clauses as introduced in chapter 2, but also technically as it seems to be able to remedy the problems the previously discussed theories have encountered. Nevertheless, the theory’s great flexibility, which is responsible for much of the heavy lifting with regard to the remedy of the different problems, leads us to the question of whether it does not overgenerate and, thus, allow for readings of counteridenticals to exist which are actually infeasible.

The first indication that this apprehension might be true can be derived from Kocurek’s assumption of the copula clause in the antecedent’s being equative. Reconsider sentence (67), whose semantics Kocurek suggested to be $(I^1 = you^2) > Bet(I^1, that\ horse^3)$ and paraphrased as ‘If a C_1 -counterpart of the speaker is a C_2 -counterpart of the listener, that counterpart will bet on every C_3 -counterpart of that horse.’ The criticism I have with respect to this analysis can be formulated in two ways: Speaking in terms of the verbal paraphrase, it does not seem self-evident what exactly the term ‘that’ as in ‘that counterpart will bet on every C_3 -counterpart of that horse’ refers to. Likely, it is meant to refer to the mutual counterpart of the speaker (under the relation C_1) and the addressee (under the relation C_2). Instead of Kocurek’s giving justice to ‘that’ counterpart’s similarity to the addressee, however, the analysis only makes explicit its connection with the speaker-counterpart. Speaking in terms of the formal analysis, this criticism can be spelled out as follows: It does not seem self-evident why the reading $(I^1 = you^2) > Bet(you^2, that\ horse^3)$, i.e. *If I were you, you would bet on that horse*, which is clearly not intended by the speaker, would be predicted to be false by Kocurek’s theory. After all, you^2 is identified by means of an equivalence relation “=” with I^1 according to his analysis of the antecedent clause, which should predict the truth of this analysis and, by implication, the validity of the sentence’s paraphrase. Intuitively speaking, ‘that’ should refer to a counterpart of the speaker which is defined such that it takes into account the truth of both of the aforementioned counterpart relations. Hence, it should have an interpretation similar to the following: $I^3 \mid I^3 = I^1 \circ you^2$, where \circ relates I^1 and you^2 in some way (e.g. for (67): C_3 is a counterpart relation made up of the belief-counterpart relation with respect to the speaker and the predicament-counterpart relation with respect to the addressee).

A further indication of Kocurek’s theory’s being too flexible becomes evident when reconsidering the sentences in (68). Whereas the semantics of (68-a) are comprehensible – neglecting the point of criticism just discussed –, it is not self-evident why the last term’s counterpart in (68-b) should be derived under the relation “with my beliefs and preferences,” too. It is understandable that Kocurek wants to match the indices of the first person pronouns in the sentence’s consequent in order to achieve a reflexive interpretation of the indexicals. Nevertheless, there does not seem to be a logical rationale as to why the last pronoun should not rather be derived via a candidate-counterpart relation, like in (68-a), which, intuitively, would capture the term’s sense much more adequately. This

discussion shows that there may exist different opinions as to how the different terms in a counterfactual should be modified, an observation which leads me to a further point of criticism with respect to Kocurek’s theory: The guidelines for assigning counterpart relations seem to be too flexible in proportion to the meaning the difference/similarity of indices plays for Kocurek’s analysis. To support this claim, let us reconsider the sentence *If I were Michelle Obama, I’d live in the White House*. We have argued that, in addition to the complete identification of the speaker with Michelle Obama, also the substitution of the counterpart index 1 by means of any predicament-relation renders the correct interpretation of the conditional. Nevertheless, let us further recall in this regard that Kocurek’s theory states that “a separate index [should be assigned] to each distinct way of filling in the blanks” of the modifying clauses ‘as ___’, ‘being ___’, etc. (Kocurek 2016: 24). With this in mind, consider the following paraphrases:

- (72) If I were Michelle Obama, I’d live in the White House.
- a. If I (in my predicament) were Michelle Obama (in her predicament), I (in my predicament)’d live in the White House.
 - b. If I (in my predicament) were Michelle Obama (as the wife of the current president of the US), I (in my predicament)’d live in the White House.
 - c. If I (in my predicament) were Michelle Obama (as the wife of the current president of the US), I (as the wife of the current president of the US)’d live in the White House.

It can be argued that different people might assign any of the above, or even yet different modifying clauses to the terms in trying to capture the sentence’s (72) correct meaning. At least the last two parses, albeit, would generate different counterpart indices for the terms in the counterfactual’s antecedent, which according to Kocurek’s algorithm, prevents an ‘imagine’ reading from arising. Nevertheless, this does not seem like a prediction we would want to make.

Lastly, Kocurek’s basic assumption – the fact that the context may provide more than one counterpart for a given individual – can also be argued to be dubitable. Consider the following argument, put forward by Arregui (2007): If it were the case that the availability of multiple counterpart relations for an individual is responsible for the difference between the sentences in (68), the same should hold for other individuals/pronouns apart from the speaker/I as well (cf. 36; 32). Nevertheless, this is not what we find:

- (73) *If Peter^{1_i} were Sarah^{2_j}, he^{1_k}’d kiss him^{3_i}.
 (same index and reference pattern as in (68-a))

According to Kocurek’s theory, it should be possible to identify Peter as two distinct individuals in the counterfactual *p*-worlds by means of two different counterpart relations,

yielding the desired interpretation (cf. (68-a)). The fact that such principle B violations are not valid for third person (and arguably second person pronouns (cf. Arregui 2007: 32)) in a counterfactual's antecedent, however, motivates us to conclude with Arregui that each context only provides one counterpart for a given individual.

All things considered, we have seen that Kocurek's theory seemed very promising at a first glance, but – like the previous counterpart theories – it also incorporates assumptions which seem questionable in light of the data from chapter 2. The problems Kocurek's theory generates are not easily remedied since the assumption of an asymmetric copular clause in the counterfactuals' antecedents, for example, would prevent the two relevant counterparts to generate a counterpart at other possible worlds which they vicariously share. Furthermore, the flexibility of the assignment of counterpart relations as well as the availability of several counterpart relations are crucial for Kocurek's proposal, but have been argued to lead to difficulties, as well.

Given that all of the counterpart theories discussed have been argued not to be able to capture the correct semantics of counterfactual conditionals in spite of the different frameworks they were developed in as well as the different degrees of freedom they incorporated, I will not pursue the counterpart approach to *If I were X*-conditionals further at this point.

6 An Analysis of Counteridenticals in Terms of Dream Reports

It has been shown that a straightforward combination of the semantics of identity statements and counterfactuality does not lead to the desired truth-conditions of *If I were X*-conditionals (cf. chapter 5.1). Furthermore, the previous chapter has gone at length to show that – so far – none of the counterpart theories of counteridenticals are able to remedy the variety of problems connected to *If I were X*-conditionals that were raised in this thesis: Whereas Lakoff (1996) and Malamud (2006) were shown to be too restrictive, Kocurek (2016) turned out to be too permissive.

For my own proposal of counteridentical meaning I have opted to take a different route, therefore: I argue that counteridenticals bear semantic as well as syntactic similarities with regard to dream reports, which motivates me to align the semantics of *If I were X*-conditionals with that of dream reports, as well.

6.1 Parallels Between Counteridenticals and Dream Reports

Let us begin by explicating the connection between the two different areas of linguistic research. The first similarity between dream reports and counteridenticals consists in the fact that they both allow for sequences to occur that would not be allowed as independent matrix clauses:

- (74) a. I dreamed I was Brigitte Bardot and I kissed me.
b. If I were you, I'd kiss me.

(cf. Arregui 2007: 31)

As has been discussed in section 5.2.2, in existential contexts, sequences like those underlined in (74) constitute violations of the binding principle B, which requires that a pronoun must be unbound within its governing category (cf. Chomsky 1981). An example of such a principle B violation is provided by the following sentence: **I looked into the mirror and I saw me*. Since in both counteridenticals as well as dream reports it is possible for the two relevant occurrences of first person pronouns to be referring to different people, namely the pronoun's referent in the actual world as well as the one in the counterfactual/dream world, they seem to be able to circumvent the binding principle.

Interestingly, however, both counteridenticals as well as dream reports only allow principle B violations for first person pronouns. We have seen that the same structure is not permitted for third person pronouns in example (73), and it is commonly argued that it is not permitted for second person pronouns, either (cf. Arregui 2007: 32). The same pattern is found in dream reports.

- (75) a. (i) *If Peter_i were Sarah_j, he_k'd kiss him_i.
 (ii) *Peter_i dreamed [he_i was Brigitte Bardot_j and] he_k kissed him_i.
 b. (i) */?If you_i were me_j, you_k'd kiss you_i.
 (ii) *You_i dreamed [you_i were Brigitte Bardot_j], and you_k kissed you_i.

A further similarity between dream reports and *If I were X*-conditionals is that they enable us to comprehend clauses which seem “irremediably false” in extensional contexts (Arregui 2007: 31). Consider the following examples.

- (76) a. I dreamed I was you.
 b. If I were you, I would be happier. (ibid.)

When evaluated against the facts of the actual world, the identification of two inherently different individuals seems clearly infelicitous. Nevertheless, in the case of dream reports and counteridenticals we can easily make sense of such a relation, since we derive from their structures that instead of consulting our knowledge of the actual world we are to imagine worlds which differ from ours with regard to some contextually relevant presuppositions, here: the identity of you and me.

In particular with regard to ‘imagine’ *If I were X*-conditionals, we realize a further similarity to dream reports: Both kinds of counterfactual identification of the speaker with another person *X* prompts the assignment of the entire set of (contextually relevant) properties defining that person *X* to the speaker on the part of the listener. If, in such a situation, the speaker wants to change any of *X*’s properties which undergo the re-ascription process, he has to make the change explicit. Consider, for instance, the following examples:

- (77) a. I dreamed I was you. But you lived in New York and had a great apartment.
 (Arregui 2007: 36)
 b. If I were Michelle Obama, I’d live in the White House, but (unlike her,) I’d love it.

We see that in both scenarios, the copular clause raises certain expectations with regard to what it means for the dreamer to be the listener in the dream worlds, or the utterer of the ‘imagine’ counteridentical to be Michelle Obama in the counterfactual worlds on the part of the addressee. If the speaker assumes a deviation from this set of properties without explicating it in the utterance, the listener is expected to object:

- (78) a. A: Yesterday, I dreamed I was you. I lived in New York and I had a great apartment . . .
 B: I don’t think it was me that dreamed that you were. I don’t live in New York and my apartment is pretty crappy.

- b. A: If I were Michelle Obama, I'd live in the White House and I'd love it.
 B: Wait ... I always thought Michelle Obama hates living there.

Somewhat less obvious seems the same requirement on the reassignment of properties in the 'advice' case, since here we only assume a partial identification process in the first place. In order to see that 'advice' counteridenticals nevertheless behave in the same way, consider the following example:

- (79) CONTEXT: Mary is at the mall with her friend Kate, who wants to buy a new dress. There are three dresses at choice: a blue, a red, and a green one, which is also the order in which Mary likes them best. Kate is asking Mary about which dress to buy.
- a. Mary: If \bar{I} ^F were you, I'd buy the blue dress. But I know that you don't like the color blue, so I guess I'd pick the red one. [which is the one Mary likes second best]

Here, the 'adjustment' of properties goes the opposite direction, since in an 'advice' counteridentical, the speaker is expected to make a statement assuming to be in someone else's external position but keeping all of his internal properties in place. In order to be able to give a useful piece of advice, however, we see that the speaker has to adjust his internal properties if the context demands so. Just like in the case of dream reports and 'imagine' counteridenticals, such a change in the set of internal properties the speaker projects into the addressee's situation in 'advice' counteridenticals has to be made explicit. One strategy has been illustrated in (79). We have seen a different strategy in the example in (53), restated here as (80):

- (80) a. If I were you, I'd go skydiving tomorrow.
 b. If you were me, you wouldn't go skydiving tomorrow.

In (80-a), the speaker states what he would actually do were he in the addressee's position in this context; nevertheless, in order for his utterance to be understood as a useful piece of advice, he has to take into account the addressee's fear of heights (80-b). This difference between the two utterances becomes evident by means of the shift in focus.

Yet another parallel between *If I were X*-conditionals and dream reports is the reference constraint which the pronouns of both constructions seem to underlie. Percus and Sauerland (2003) observe that in a dream report, which involves two pronouns (e.g. John dreamed he married his grand-daughter), the first pronoun can never refer to the actual self if the second pronoun refers to the dream-self. To illustrate this claim, consider the following example:

- (81) John dreamed that he was marrying his grand-daughter.
- a. In his dream, the dream-self marries the dream-self's grand-daughter.
 - b. In his dream, the dream-self marries John's grand-daughter.
 - c. In his dream, John marries John's grand-daughter.
 - d. *In his dream, John marries the dream-self's grand-daughter.
- (cf. *ibid.*: 4)

Percus and Sauerland argue that this restraint in reference is due to a syntactic constraint on pronoun movement (Oneiric Reference Constraint (ORC)), which they propose to be responsible for the different readings of sentences like (81). Interestingly, we find a similar restraint on pronoun reference in counteridenticals, as well.

- (82) If I were you, I'd encourage my son to play with my daughter.
- a. If I_i were you $_j$, I_k 'd encourage my $_j$ son to play with my $_j$ daughter.
 - b. If I_i were you $_j$, I_k 'd encourage my $_j$ son to play with my $_i$ daughter.
 - c. If I_i were you $_j$, I_k 'd encourage my $_i$ son to play with my $_i$ daughter.
 - d. ?/* If I_i were you $_j$, I_k 'd encourage my $_i$ son to play with my $_j$ daughter.

It has been shown in chapter 2 that the subject pronoun in the consequent of *If I were X*-conditionals consists of a combination of properties of the speaker and the addressee and that the degree of identification between the speaker and the addressee determines the kind of counteridentical we are dealing with. Evidently, this pronoun can never refer to the actual self of the speaker, therefore. In order to see if the ORC also holds for counteridenticals, we thus have to look for two pronouns whose referents can be chosen freely between the actual self of the speaker and the person he counterfactually identifies with. Such a structure is represented by (82). We see that – like in the case of dream-reports – the reading in which the actual speaker's son is to be encouraged to play with the counterfactually imagined daughter is infeasible, or at least the least accessible.

Given the numerous parallels, which exist between dream-reports and counteridenticals, I deem it only to be logical that their semantic analyses should be similar, as well. Hence, in the next section, let us take a look at existing proposals of dream-report meaning, and let us discuss how these analyses can help us to understand the semantics of *If I were X*-conditionals.

6.2 Discussion of Existing Proposals of Dream Reports

Since we have just discussed the parallel in reference constraint as proposed by Percus and Sauerland (2003), let us start by sketching their proposal.

6.2.1 Percus and Sauerland (2003)

Percus and Sauerland (2003) base their analysis of dream reports on a possible worlds analysis of attitudes as proposed by Hintikka (1969) and Lewis (1979) (cf. 6). Hence, they assume that the range of one's dreams can be conceptualized as a range of possible worlds, within each of which the dreamer unequivocally situates himself.¹⁷ Due to this assumption, their analysis of 'dream' can be characterized as "believe-while-asleep" (Arregui 2007: 38). Furthermore, they adopt Lewis' (1973) idea that each individual can inhabit exactly one world, meaning that the dreamer experiences each dream world vicariously through a designated counterpart (cf. chapter 5.2.1).

On the basis of possible world analyses of attitude contexts, as well as Lewis' counterpart theory, Percus and Sauerland (2003) put forward the following meaning for the two possible referents of the same pronoun: In cases where the dream-self is picked out by the pronoun, their analysis yields the center of the relevant possible world, i.e. the individual with whom John identifies in the designated dream worlds. In cases where it is the actual dreamer who is referred to by means of the pronoun, their analysis yields a certain individual concept which holds of that person at a dream world w' , $k(w')_{\langle s, e \rangle}$, representing the counterpart of the actual person at that world. Given Lewis' definition of counterparts, such an individual concept picks out exactly that individual at w' which we intuitively might think of as the dreamer (cf. 7). This analysis can be schematized as follows:

- (83) In John's dream, he gets promoted.
- a. In John's dream, the dream-self gets promoted.
 - (i) Given any pair $\langle y, w \rangle$
 - where w is a world compatible with John's dream
 - and y is the individual in w who John identifies as himself,
 - y gets promoted in w .
 - b. In John's dream, John gets promoted.
 - (i) Given any w
 - where w is a world compatible with John's dream,
 - $J(w)$ gets promoted. (ibid.)

According to Percus and Sauerland (2003), the ORC supports the idea that these two different interpretations of the pronoun *he* come about via a conceptualization of the pronouns as different kinds: Whereas they suggest to analyze those pronouns picking out the counterpart of John at his dream worlds in correspondence with their usual interpretation as variables (cf. Heim and Kratzer 1998), they posit that those pronouns

¹⁷The assumption of not being in doubt about which individual to identify with is contra Lewis (1979).

picking out John’s dream-self behave similar to relative pronouns: they do not receive an interpretation *in situ* (marked by an asterisk) but move to the left periphery of the complement clause, which triggers a predicate abstraction over the trace they leave behind (cf. 7f). Since Percus and Sauerland assume the denotation of ‘dream’ to be similar to that of attitude verbs, i.e. they assume that ‘dream’ quantifies over centered worlds and takes a property (the meaning of the complement clause) as an input (cf. (84)), such a movement leads to an identification of the moved pronoun with the center of worlds that are compatible with John’s dream worlds, i.e. y .

$$(84) \quad \llbracket \text{dream} \rrbracket^g = \lambda P. \lambda x. \lambda w. \text{For all } \langle y, w' \rangle \text{ in } \text{DREAM}_{x,w}, P(y)(w') = 1.$$

($\text{DREAM}_{x,w}$ stands for the set of pairs $\langle y, w' \rangle$ such that w' is a world compatible with x ’s dream in w , and y is the individual in w' who x , in w , identifies as himself.)

(Percus and Sauerland 2003: 8)

The logical form of the string “(John) dreamed that he_{dream-self} was marrying his_{actual-self} grand-daughter.” under this proposal is the following:

$$(85) \quad \begin{aligned} & \text{(John) dreamed that } \text{he}_{\text{dream-self}} \text{ was marrying } \text{his}_{\text{actual-self}} \text{ grand-daughter.} \\ & \text{a. } \text{dream} [\text{he}^* \lambda_3 [\lambda w_1 [_{VP} w_1 t_3 \text{ was marrying } [\text{his}_2 w_1] \text{ grand-daughter}]]] \\ & \text{b. } \lambda x. \lambda w. \forall \langle y, w' \rangle \text{ in } \text{DREAM}_{x,w}, y \text{ marries the grand-daughter of } g(2)(w') \\ & \quad \text{in } w'. \end{aligned}$$

(ibid.: 10)

In this analysis, the different colors serve to illustrate which parts of the string are involved in the analysis of which pronoun. We see that the unstarred pronoun is analyzed *in situ* as it combines with a world parameter which, due to lambda-abstraction, receives its denotation from the worlds compatible with John’s dream worlds, i.e. w' . The starred pronoun behaves as described above. The only thing left to do to yield the desired sentence meaning is to lambda-abstract over individual concepts in order to prevent $g(2)$ to be identified with the actual John, but rather with the individual concept of John at w' , $J(w')$. All in all, Percus and Sauerland thus posit the following semantics for the string in (85), which then takes the dreamer – John – as an input.

$$(86) \quad \lambda k. \lambda w. \forall \langle y, w' \rangle \text{ in } \text{DREAM}_{k(w),w}, y \text{ marries the grand-daughter of } k(w') \text{ in } w'. \quad (\text{ibid.: 11})$$

The ORC now excludes all those structures by means of a concept which Percus and Sauerland (2003) call “superiority” where a starred pronoun *pro** would have to move across an unstarred pronoun which a) asymmetrically c-commands it and which b) shares

the same features pro^* has (cf. 14). Such a structure can be represented by means of the following logical form:

(87)

$$* \dots \lambda_i^c [_{VP} t_i^c \dots [_{V'} \text{dream} [\text{pro}^{*} \left[\lambda_j [\dots \mathbf{pro}_i[\text{actual self}] \dots t_j[\text{dream-self}] \dots]]]]],$$

where pro_i asymmetrically c-commands t_j and all embedded pronouns with index j (ibid.: 12)

As per Percus and Sauerland, such a logical form underlies dream reports like (81-d), and, by implication, possibly also counteridenticals like (82-d).

In spite of the appeal of Percus and Sauerland's theory in general as well as with respect to counteridenticals in particular, there has been independent evidence of the fact that we might need more than one strategy to interpret pronouns in attitude contexts *de se*, i.e. as picking out the dream-self, by Anand (2006) as well as by Ninan (2008). Anand, for example, motivates this proposal on the basis of his observation that in a *believe*-context, the blocking effect proposed by Percus and Sauerland does not hold (cf. Anand 2006: ch. 1.4.1). Since it is the aim of this section to explore the toolkit which dream reports make use of in order to understand the semantics of counteridenticals more thoroughly, let us briefly sketch one of the proposals incorporating different strategies to a *de se*-interpretation.

6.2.2 Ninan (2008)

Proponents of the existence of more than one *de se*-strategy in dream reports agree with Percus and Sauerland (2003) in that the *de se* reading can be derived via specialized binding logical forms. Nevertheless, in addition to this syntactic strategy, they further agree with Lewis (1979) in that a *de se* interpretation can also be derived by means of a designated way of understanding *de re* pronouns, namely if a pronoun is interpreted *de re* via a concept which traces individuals across worlds by means of their being the entities the individual under discussion identifies with (cf. Arregui 2007: 37; Lewis (1979)). On this basis, Ninan (2008), for example, assumes the following analysis for dream reports: the dream-self is picked out *de se* by means of specialized binding, while reference to the actual dreamer is achieved via a suitable *de se-as-de re* strategy (cf. Pearson 2015: 10). Such a *de se-as-de re* strategy can, for example, be modeled in the subsequent way:

- (88) SELF-based concept
- a. C is a SELF-based concept of x for x iff for every world-individual pair $\langle w', y \rangle$ that is in the domain of C , SELF holds between y and $C(w', y)$,
- where $\text{SELF}(x, y, w) = 1$ iff (i) x is sentient in w and (ii) $x = y$.
- (ibid.: 8)

It becomes evident that the SELF-based concept does exactly what was suggested before; it ensures that a *de se*-reading arises as a special case of *de re*, since the property used to identify the relevant individual's counterpart at each world is defined as their being the entities the individual identifies with (i.e. for all centered worlds, the SELF-relation will pick out the doxastic center at that world, $C_{\text{SELF}}(w', y) = y$) (cf. Pearson 2015: 8).

In order to combine the two *de se*-strategies into a unified proposal of the meaning of imagination (or dream) reports, Ninan assumes that each act of imagining/dreaming is relative to one's beliefs: If I *imagine/dream* to be a different person, my *beliefs* about who I am remain unchanged (cf. *ibid.*: 10). This leads to a semantics of attitude verbs that assumes the quantification over pairs of centered worlds instead of over only one centered world:

$$(89) \quad \llbracket \text{imagine} \rrbracket^{c,g} = \lambda P : P \in D_{\langle e, \langle s, t \rangle \rangle}. \lambda x : x \in D_e. \lambda w : w \in D_s. \forall \langle \langle w', y \rangle, \langle w'', z \rangle \rangle \in \text{Imagine}_{x,w}, P(z)(w''),$$

where $\text{Imagine}_{x,w} = \{ \langle \langle w', y \rangle, \langle w'', z \rangle \rangle : \text{It is compatible with what } x \text{ believes in } w \text{ for } x \text{ to be } y \text{ in } w', \text{ and it is compatible with what } x \text{ imagines in } w \text{ relative to } \langle w', y \rangle \text{ for } x \text{ to be } z \text{ in } w''. \}$

(Pearson 2015: 15; based on Ninan 2008: 44f)

With these semantics in place, let us now see how Ninan (2008) deals with the analysis of imagination/dream reports in practice:

- (90) George: I'm imagining (that I am Brigitte Bardot and) I_{B.B} am kissing me_S
- a. 'I'm imagining (that I am Brigitte Bardot and) I am kissing me.' is true in w with respect to a SELF-concept of George for George in w iff for all $\langle \langle w', y \rangle, \langle w'', z \rangle \rangle : \langle \langle w', y \rangle, \langle w'', z \rangle \rangle \in \text{Imagine}_{\text{George},w}, z [= \text{Brigitte Bardot}]$ is kissing $C_{\text{SELF}}(w', y) [= y = \text{George}]$ in w'' .
- \Rightarrow *George is imagining that the individual that he is imagining himself to be (Brigitte Bardot) is kissing the individual that he believes himself to be (in the ordinary course of things, George himself).*

(based on Pearson 2015: 15/Ninan 2008)

We see that z picks out the dream-self via a special *de se*-binding relation, similar to that proposed by Percus and Sauerland (2003), whereas $C(w', y)$ under the SELF-relation picks out the person George identifies with in his beliefs, i.e. his own counterpart.

6.3 Direct Transfer to Counteridenticals?

In section 6.1, several parallels have been made out between the structures of dream reports and counteridenticals. Furthermore, the discussion of the meaning of dream reports

has shown that the two areas of linguistic research also seem related in a conceptual way: in both cases, the speaker counterfactually imagines some person (often himself) to be someone else; what is more, for both cases it has been suggested that the worlds the speaker counterfactually imagines are anchored to his belief-worlds. What we have just seen at play in Ninan (2008), researchers like Pelletier (2004) and Recanati (2000) have suggested for *If I were X*-conditionals: they posit that the counterfactual *p*-worlds against which the claim in the counterfactual's consequent is tested are generated on the basis of a notion of the speaker's "pretence-cum-betrayal" (Pelletier 2004: 151), whereby the speaker does not lose track of the fact that he is only playing a part in *If I were X*-scenario, but crucially still believes to be himself (cf. *ibid.*). All things considered, taking over elements from the analyses of dream reports seems promising with regard to the analysis of counterfactual conditionals, therefore.

Nevertheless, if one tries to align the analysis of counterfactuals with that of dream report, a technical difficulty arises: the specialized *de se*-binding strategy, which both of the discussed proposals of dream reports crucially rely on, does not seem to be feasible for the case of *If I were X*-conditionals. This infeasibility is grounded in the fact that the starred pronoun is dependent on the left periphery of the embedded CP as the landing site for its lambda abstractor, since it wants to be identified with the center of the speaker's doxastic worlds (cf. chapter 6.2). In the case of counterfactuals, or conditionals in general, albeit, no such landing site is available.

A remedy for this problem might be derived from Moltmann (2003), who suggests that propositions should be redefined as attitudinal objects. As per Moltmann, attitude verbs do not specify relations between agents and propositions, but rather between agents and propositional constituents (cf. 78). Since each of these constituents may be thought of in a certain way, or under a certain "mode of predication" (*ibid.*), it is not just one single attitude relation that underlies attitude reports but several, depending on the propositional content (cf. *ibid.*). In order to illustrate Moltmann's proposal, consider the following sample analysis:

- (91) John believes that Mary is happy.
 a. $R(\text{John}, \langle \text{Happy}, T_1 \rangle, \langle \text{Mary}, T_2 \rangle)$
(cf. Moltmann 2003: 94)

Here, *R* is a belief-relation which connects the agent of the sentence, John, to the propositional constituents of the embedded CP. Each of the propositional constituents is further perceived under a specific mode of presentation, represented by T_i .

Up to this point, Moltmann's analysis merely constitutes a different way of analyzing attitude reports. The part of her proposal which seems promising with respect to the problem we have faced in the analysis of counterfactuals, becomes evident when considering her analysis of simple declarative sentences.

(92) Mary is happy.

a. $\lambda x[\mathbf{R}_{(\text{ass},3)}(x, \langle \text{Happy}, T_1 \rangle, \langle \text{Mary}, T_2 \rangle)]$

b. An agent predicates, in the assertive mode, the property of being happy_{T₁} of Mary_{T₂}

(cf. Moltmann 2003: 98)

In the case of declarative sentences, we are not explicitly dealing with any kind of attitude report. Nonetheless, Malamud proposes to conceptualize them in a parallel manner: She suggests that the utterer of a declarative statement is connected to the propositional constituents making up the declarative statement by means of an assertion-relation.

With respect to counteridenticals, Moltmann’s (2003) approach can be taken to provide the means for generating a landing site for the starred pronoun: On the basis of her theory, a counteridentical may be conceptualized as an attitudinal object, which is connected to the agent by means of an ‘entertaining’ relation, whereby the entertaining relation is understood as “the most general attitudinal relation there is” (99), i.e. one which is not restricted to any specific sentence/attitude type. This suggests that counteridenticals may be paraphrased in the following way: “The speaker (counterfactually) entertains [*CP* that *if I were X, . . .*],” which clearly incorporates the desired embedded *CP*.

For now, let us disregard the technical implications a proposal combining Moltmann’s theory with that by Percus and Sauerland/Ninan would entail (i.e. figuring out which propositional constituents are at play and which manner of their combination the attitude verb should trigger), and let us simply acknowledge, based on the previous discussion, that such a combined proposal seems promising with respect to the research question of this thesis. Nevertheless, let us also acknowledge the fact that if such an analysis is to work for counteridenticals, it needs some refinement: Even though the straightforward combination of the different proposals might work for the purpose of this thesis, whose aim it is to analyze the meaning of *If I were X*-conditionals, it seems evident that it would not work for counteridenticals like *If Peter were Susan, he’d ask that guy out*, where the center of the counterfactual *p*-worlds is not to be identified with the person whom the speaker counterfactually imagines to be himself, but the person Peter is counterfactually imagined to be. Hence, the embedding of such if-clauses into the matrix clause “The speaker (counterfactually) entertains . . .” does not yield the desired result. Furthermore, it is not obvious how such an approach should account for the difference between the ‘advice’ and the ‘imagine’ reading of counteridenticals.

All things considered, the approach’s proposed ability of capturing the meaning of *If I were X*-conditionals notwithstanding, it has become evident that it would yield only a partial solution to the general problem of counteridentical meaning. Hence, this thesis does not offer it as a final solution, but instead suggests a variation thereof as an answer to its research question.

6.4 Proposal

The proposal I would like to put forward makes two refinements to the approach discussed in the preceding subsection, i.e. the combination of Moltmann’s (2003) idea to redefine propositions as attitudinal objects, and a suitable analysis of dream reports (e.g. Percus and Sauerland (2003) or Ninan (2008)). Even though I will not be able to provide the technical details of the proposals in this thesis, note that such an analysis of counteridenticals presupposes both the antecedent- as well as the consequent worlds to be centered worlds. Hence, their intersection (which ensures the conditionality requirement) will consist of centered worlds, as well.

On the basis of this general assumption, I propose to reinterpret Moltmann’s idea that any utterance can be interpreted as an attitudinal object in a way such that the utterance’s propositional content is no longer bound by the agent. Instead I suggest an underlying structure which attitudinally relates the agent to a property he entertains of an entity a (cf. (93)). Since the embedded clause (here: the counteridentical) only attaches below this matrix clause, the lambda abstractor responsible for the interpretation of the starred pronoun, which yields the counterfactual p -world counterpart of a according to the dream-report proposals, can receive its information from a . In other words, I suggest that a counteridentical may be paraphrased as follows:

- (93) An agent entertains of a that $p(a) \Leftrightarrow$ An agent entertains of a : $\{ \langle a, w \rangle \mid p(a) \}$,
where $p(a)$ is a counteridentical property which holds of a in w .

By means of such a redefinition of Moltmann’s idea, a can now bind all of the relevant subject pronouns, i.e. “If I/Peter were you/Susan, I/he . . . ”

But what exactly is the meaning of the center a ? I propose that counterfactual worlds are centered around situations and, furthermore, that it is in the denotation of the center that the ‘advice’ and the ‘imagine’ reading of counteridenticals come apart. Let me briefly motivate this idea. One of the main differences which has been made out between the two readings of *If I were X*-conditionals is that in the case of ‘imagine’ counteridenticals, the proposition incorporated in the structure’s consequent already holds of the object of the identity statement, whereas in the case of ‘advice’ counteridenticals, it must not hold of either of the individuals mentioned in the antecedent (cf. chapter 2). Consider the following examples:

- (94) a. If I were Michelle Obama, I’d live in the White House.
b. If I were Michelle Obama, I’d fly to Hawaii tomorrow.

In sentence (94-a), the property of the consequent clause, i.e. “living in the White House” constitutes a subset of the properties defining Michelle Obama in the actual world. Were we to neglect the antecedent and substitute the I in the consequent clause by ‘Michelle

Obama,’ we would arrive at a true statement, “In w_0 , Michelle Obama lives in the White House,” therefore. By contrast, the property expressed by the consequent in (94-b) must not necessarily hold of either the speaker or Michelle Obama in the actual world. Assume that Michelle Obama is not flying to Hawaii tomorrow, and neither is the speaker (e.g. because he does not have the necessary money or time). Under such circumstances, the speaker may utter the advice in (94-b), but neither the statement “In w_0 , the speaker flies to Hawaii tomorrow,” nor the statement “In w_0 , Michelle Obama flies to Hawaii tomorrow” is true.

Recalling chapter 3.1, we see that in order to fulfill the conditionality restriction, it would be sufficient for the center in (94-a) to contain the situation of Michelle Obama, whereas in (94-b), the center’s identity has to be constructed from a combination of the situation of Michelle Obama and that of the speaker in order to make the claim in the sentence’s consequent true for the center. To this end I claim – not too controversially – that character is a subset of situations, an assumption, which, for example, may be buttressed by means of the following example sentences.

- (95) a. If Peter weren’t $\overline{\text{Peter}}^F$, he would come out with us tonight.
 \Rightarrow *Intended meaning: If Peter weren’t so conscientious, he would come out with us tonight.*
- b. Boss: Yesterday, Peter ordered his colleagues about as if he were me.
 \Rightarrow *Intended meaning: Yesterday, Peter ordered his colleagues about as if he were the boss.*

Thus, I propose the following meaning of the center a for counteridenticals: As a default, the center contains the situation of the person whom the subject of the identity statement identifies with. In ‘advice’ cases, the center is further adapted by means of the contextually relevant internal properties of the speaker. Underlying both of these identification processes is an asymmetric identification process of the subject of the identity clause with its object as suggested per Percus and Sharvit (2014) (cf. chapter 4.2.2.2). The degree of identification varies across the two readings.

Lastly, let us see how this proposal accounts for the multiplicity of pronoun reference in the consequent clause of counteridenticals, which is the only aspect we have not yet considered in the analysis. The above paragraphs have laid out the theory’s strategy to refer to the counterfactual-self of the identity relation’s subject; the reference to the subject’s actual-self can now be accounted for in several ways: Either we may choose a strategy similar to Percus and Sauerland (2003), i.e. assume that there is an independent lambda-abstraction over world-variables, which the pronouns are combined with *in situ*, or we may choose a strategy similar to Ninan (2008), i.e. assume that there exist designated concepts, which unequivocally pick out the actual individuals in the speaker’s belief-worlds;

nevertheless, we could also choose a different syntactic approach and conceive of them as donkey anaphora, or assume that the subject of the antecedent clause moves to the topic position of the sentence from where it can bind the relevant pronouns directly, as has already been suggested in chapter 5.1. Given that the technical details of the proposals are yet to be explicated, I will not give preference to any of these strategies at this point.

In summary, we conclude that the proposed theory of counterfactual meaning seems promising for several reasons: It is able to explain the suggested grammatic difference between the ‘advice’ and the ‘imagine’ reading of *If I were X*-conditionals, it takes into account the structure’s syntactic as well as semantic/pragmatic parallels to dream reports, and it seems to be able to account for the problems the existing proposals of counterfactual meaning revealed. Furthermore, the redefinition of counterfactuals as attitudinal objects also seems promising for a reason not predicted by any of the other proposals – it may give us the technical toolkit to explain why some of the indexicals in *If X were Y*-conditionals are anchored to the speaker whereas others are anchored to the counterfactual counterpart of the identity statement’s subject. To illustrate the fact that two deictic centers may be involved in the analysis of counterfactuals, consider the following example sentences, where the relevant deictic center has been made explicit by means of subscripts.

- (96) a. If I were Mary, I wouldn’t be dating that horrid guy_[attitude of speaker].
 b. If I were you, I wouldn’t be here_[speaker] right now_[speaker].
- (97) CONTEXT: Looking at a picture of Mary, who is at the beach in Spain.
 a. If I were Mary, I would taste all of the local_[Mary] food.
 b. If I were Mary, I’d jump into the sea in front of me_[Mary].

To the best of my knowledge, this duality of deixis has not been noted in the literature before, and is not easily accounted for by any of the existing accounts of counterfactual meaning. The theory proposed in this thesis, however, is able to predict such a duality by implying the existence of a global/matrix (i.e. speaker) and local/embedded (i.e. counterfactual) deixis.

In spite of the advantages of the suggested theory, it incorporates at least two aspects which one might be critical of. First of all, it has been made explicit several times throughout this thesis that *If I were X*-conditionals are, first and foremost, counterfactual conditionals. Hence, we concluded that their analysis should be compatible with the standard theories of counterfactual meaning (cf. chapter 3). From the given proposal it does not seem self-evident, albeit, how such a compatibility could be achieved without effort, since it assumes the quantification over centered attitude worlds instead of over possible worlds in general, for example. Finding the answer to this question will be left to future research.

The second aspect of the theory one might be doubtful of is that while the ‘advice’ use of counteridenticals seems to be the more natural one, the theory proposes its center to be more complex than that of the ‘imagine’ use, which might suggest – contrary to our intuition – that the parse of ‘advice’ counteridenticals takes more effort. A possible explanation for this caveat has been suggested in chapter 2.1, i.e. that – in fact – making/parsing a claim of the less complex center in ‘imagine’ readings may be more difficult for the speaker/listener, since more properties have to be counterfactually reassigned in the thought process.

7 Conclusion and Outlook

This thesis focused on the semantic analysis of *If I were X*-conditionals, a subclass of counterfactual conditionals, whose antecedent clauses identify two inherently incompatible entities with each other. After having argued that the antecedent clause *If I were X* is grammatically ambiguous between an ‘advice’ and an ‘imagine’ reading, the two of which may be distinguished on the basis of the different meanings the pronouns in their consequent clauses possess, in chapter 2, this thesis went on to answer the following research question: *How do we achieve the different denotations of the pronouns in the ‘advice’ and the ‘imagine’ use of counterfactuals, and, thereby, the different readings?*

Since, first and foremost, counterfactuals constitute a subclass of counterfactual conditional sentences, we began our analysis by sketching the two prevalent theories of counterfactual meaning to this day, i.e. Iatridou’s (2000) past-as-unreal approach and Ippolito’s (2013) past-as-past approach, and discussing them in light of our findings from chapter 2. We argued that both theories cannot trivially account for the ambiguity of *If I were X*-conditionals but that they are reliant on the denotation given to the identity statement *I be-PAST X* for their analysis. It was argued that the case of counterfactuals might leave us biased towards Iatridou’s theory rather than to Ippolito’s, since the historical accessibility relation the latter assumes is not philosophically trivial for the case of *If I were X*-conditionals. Both theories had trouble accounting for the meaning of ‘advice’ counterfactuals, which have been shown to incorporate two dissimilarities with respect to the actual world instead of the commonly assumed one dissimilarity which renders a conditional contrary to fact. This motivated us to take a closer look at the meaning of identity statements in the next chapter.

Chapter 4 then laid down the second part of the theoretical background needed for a compositional analysis of *counterfactuals*, i.e. the meaning of identity statements and their respective copula in different contexts. This thesis argues that there are two different notions of identity statements which underlie the analysis of counterfactuals: asymmetric identity statements which set up the identity statement in the clause’s antecedent, and a set of covert symmetric identity statements involved in the pronoun resolution across the *p*- and *q*-worlds. In particular, such an analysis thus favors asymmetric, predicative copulas as per Percus and Sharvit (2014) over symmetric, equative ones for the analysis of the antecedent clause, a claim which is not uncontroversial in the literature, but which has been further fortified in the discussion of the existing proposals of counterfactuals in chapter 5.2.

Before we discussed the existing analyses of counterfactual meaning, all of which can be subsumed under the general term ‘counterpart theories,’ since they constitute variations of Lewis’ (1973) counterpart theory, it was proven that a straightforward combination of the theories of counterfactual meaning and of identity statements is not

successful (cf. chapter 5.1). Hence, the technical analyses of the two subcomponents of counteridenticals were shown to provide the underlying technical understanding for the proposals to come, which needed to be enriched by different further notions. We showed that analyses which only take into consideration the notion of cross-world rigidity, as discussed by Lewis, were not able to successfully capture the set of relevant data: while the theories of Lakoff (1996) and Malamud (2006) were too restricted, Kocurek’s (2016) theory turned out to still be too permissive and, thus, to overgenerate. These insights paved the way for a new analysis of counteridenticals.

The analysis proposed in chapter 6 of this thesis refrains from following the approach of counterpart theories and instead opts to exploit the many parallels *If I were X*-conditionals show with respect to dream reports. A discussion of different semantic theories of dream reports revealed that counteridenticals have to be redefined as propositional attitudes (cf. Moltmann 2003) in order to align their meaning with that of dream reports. Furthermore, it was suggested to refine Moltmann’s theory to the effect that an agent assigns a property to the center of counterfactual worlds, which may or may not be associated with the agent himself. This way, we did not limit our analysis to *If I were X*-clauses but kept its potential to account for the entire class of counteridenticals; furthermore, by this means, we were able to provide an explanation of the duality of deixis which plays a role in the analysis of counteridenticals. Lastly, we proposed that the way the construction of the center of the counterfactual worlds is responsible for the distinction between the conditional’s ‘advice’ and the ‘imagine’ reading.

If the connection between dream reports and counteridenticals is correct, several questions present themselves as topics for future research. First and foremost, it seems necessary to investigate how the theory proposed in this thesis is compatible with the standard semantics of counterfactuals. To this end, the question of whether the proposal at hand needs to be refined in a way such that it quantifies over possible worlds, or if we might learn something from the case of counteridenticals about the meaning of counterfactuals in general, i.e. that they possibly quantify over possible worlds, too, arises.

On a different note, it seems fruitful to investigate the contribution focus has on the analysis of counteridenticals. The prediction is that focus may shift the direction of identification in the antecedent clause (98), or may vary the set of properties the speaker takes along with him when giving advice (99).

(98) CONTEXT: We broke your/my mom’s favorite vase while playing and know that she is going to be very mad about this. As we stand in the hallway, trying to figure out what to do, we hear her car approach. You ask me what I {would do in your shoes}/{am going to do now} and I respond:

- a. If I_i were you $_j$, I_i ’d stay down here and explain to her what happened. [I’m sure she’ll understand.]

- b. If I_i were $\overline{y\bar{o}u}_j^F$, I_j 'd get upstairs this second! [You always run away from responsibility. I, however, will just stay down here and explain to her what happened. I'm sure she'll understand.]
- (99) CONTEXT: You are afraid of heights and have been invited to go skydiving tomorrow. I am not afraid of heights at all and love doing adventurous things. You have asked me what I think you should do.]
- a. If \bar{I} were you, I'd go skydiving tomorrow. (I'd love that!)
- b. If I were $\overline{y\bar{o}u}^F$, I wouldn't go skydiving tomorrow. (You'd hate it!)

A similar pattern of taking into consideration the situation of the addressee when giving advice has been observed in what I call 'Rewe'-conditionals. In these cases, when asked the same question, i.e. "Where can I buy some really good cheese?," the speaker may adapt his answer according to the financial situation of the addressee in order to give contextually optimal advice, even though he might have a clear ranking of places to buy good cheese at in mind and should – intuitively – give the same answer to all interlocutors. Knowing that Peter has a lot of money, the speaker might answer "If I were you, I'd go to Dean & DeLuca to buy cheese," whereas knowing that Mary only has little money, he can say "If I were you, I'd go to Rewe to buy cheese," assuming that the speaker believes that Rewe has really good cheese in a certain price range.

A different aspect which seems worth investigating in the case of counteridenticals is the contribution of tense and aspect. In this context, it, for example, seems interesting why the PRES-IND in the consequent clause in Korean forces an 'advice' reading to arise. What is more, it has been suggested that the choice of progressive tense in the consequent clause triggers an 'imagine' reading, which is not trivially predicted by any of the analyses:

- (100) a. If I were Mary, I'd be swimming right now.
 b. If I were Mary, I'd swim right now.

Furthermore, one might investigate why the structure (*If I were*) *in your shoes, ...* is not a 'competitor' in languages like English, German or French under a variant of a principle like Maximize Presupposition (Heim 1991), whereas it is in Polish and LIBRAS.

The inquiries of these questions may help to paint a clearer picture of the meaning of counteridenticals in particular, and that of counterfactuals in general, but will have to be left to future research.

References

- Ahn, Dorothy. Personal interview. 4 Apr. 2016.
- Anand, Pranav. *De De Se*. Unpublished PhD dissertation, MIT, 2006.
- Anderson, Alan. "A note on subjunctive and counterfactual conditionals." *Analysis* 12 (1951): 35-38.
- Arregui, Ana. "Being Me, Being You: Pronoun Puzzles in Modal Contexts." *Proceedings of Sinn und Bedeutung*. 11. Ed. E. Puig-Waldmüller. Barcelona: Universitat Pompeu Fabra, 2007. 31-45.
- Bondaruk, Anna. *Copular Clauses in English and Polish. Structure, Derivation and Interpretation*. Lublin: Wydawnictwo KUL, 2013.
- Cable, Seth. "The Semantics of Conditionals." Class Handout. Formal Semantics Ling 620. UMass Amherst, Amherst. Spring 2013. *umass.edu*. Web. 5 Apr. 2016. <<http://people.umass.edu/scable/LING620-SP13/Handouts/>>.
- , "Introduction to (Some of) The Puzzles Surrounding the De Dicto / De Re / De Se Ambiguity." Class Handout. Semantics/Philosophy of Language Seminar Ling 753/Phil746. UMass Amherst, Amherst. Spring 2015. *umass.edu*. Web. 7 Apr. 2016. <<http://perezcarballo.org/ling753-phil746/files/handout1.pdf>>
- Cartwright, Richard. "On Singular Propositions." *Meaning and Reference*. Ed. A.A. Kazmi. Calgary: Calgary UP, 1998.
- Chierchia, Gennaro. *Dynamics of Meaning*. Chicago: Chicago UP, 1995.
- Chomsky, Noam. *Lectures on Government and Binding*. Dordrecht: Foris, 1981.
- Chrisholm, Roderick. "Identity through Possible Worlds: Some Questions." *Nous* 1 (1976): 1-8.
- Citko, Barbara. "Small clauses reconsidered: Not so small and not all alike." *Lingua* 118 (2008): 261-295.
- Csipak, Eva, and Regine Eckardt. "What are Conditionals?" Class Handout. Hauptseminar: Conditionals. Georg-August-Universität Göttingen, Göttingen. Summer 2013.
- Cumming, Samuel. "Variabilism." *The Philosophical Review* 117 (2008): 525-554.
- Davidson, Kathryn. *Quotation, demonstration, and iconicity*. Manuscript Draft. 1 Sep. 2014. Yale University.
- Declerck, Renaat, and Susan Reed. *Conditionals: A Comprehensive Empirical Analysis*. Berlin / New York: de Gruyter, 2001.
- Fauconnier, Gilles. *Mental Spaces*. MIT UP: Cambridge, MA, 1985.
- Føllesdal, Dagfinn. "Essentialism and Reference." *The Philosophy of W.V. Quine*. Eds. Lewis E. Hahn, and Paul A Schilpp. LaSalle: Open Court, 1986.
- Frege, Gottlob. "Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought." *From Frege to Gödel: A source book in mathematical logic*,

- 1879-1931 (9). Ed. Jean van Heijenoort. Cambridge, MA: Harvard UP, 1967. (1879): 1-82.
- Fuchs, Zuzanna. Personal interview. 4 Apr. 2016.
- Goodman, Nelson. *Fact, Fiction and Forecast*. Cambridge, MA: Harvard UP, 1984.
- Gordon, Robert M. "Sympathy, simulation, and the impartial spectator." *Ethics* 105.4 (1995): 727-742.
- Han, Chung-hye. "Comparing English and Korean Counterfactuals: the Role of Verbal Morphology and Lexical Aspect in Counterfactual Interpretation." *Proceedings of ESCOL '96*. Web. 5 Mar. 2016. <<http://www.sfu.ca/~chunghye/papers/escol96long.pdf>>
- Heim, Irene. "E-type pronouns and donkey anaphora." *Linguistics and philosophy* 13.2 (1990): 137-177.
- , "Artikel und Definitheit." *Semantics: An International Handbook of Contemporary Research*. Berlin: de Gruyter, 1991.
- Heim, Irene, and Angelika Kratzer. *Semantics in generative grammar*. Vol. 13. Oxford: Blackwell, 1998.
- Higgins, Roger Francis. *The Pseudo-cleft Construction in English*. New York: Garland, 1979.
- Hintikka, Jaako. "Semantics for propositional attitudes." *Philosophical logic*. Eds. J.W. Davies et al. Dordrecht: Reidel, 1969.
- Iatridou, Sabine. "The Grammatical Ingredients of Counterfactuality." *Linguistics Inquiry* 31.2 (2000): 231-270. *mit.edu*. Web. 5 Apr. 2016. <<http://web.mit.edu/Linguistics/people/faculty/iatridou/counterfactuality.pdf>>
- "Identity statement." *The Blackwell Dictionary of Western Philosophy*. 2009.
- Ippolito, Michela. *Subjunctive Conditionals: A Linguistic Analysis*. Cambridge, MA: MIT P, 2013.
- Johnson-Laird, Philip Nicholas. *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard UP, 1983.
- Kamp, Hans. "A Theory of Truth and Semantic Representation." *Formal Methods in the Study of Natural Language*. Eds. Groenendijk et. al. Amsterdam: Amsterdam Centre, 1981.
- Karawani, Hadil. *The Real, the Fake, and the Fake Fake in Counterfactual Conditionals, Crosslinguistically*. Dissertation, Universiteit van Amsterdam, 2014.
- Karawani, Hadil, and Hedde Zeijlstra. "The semantic contribution of the past morpheme in Palestinian counterfactuals." *Journal of Portuguese Linguistics* 12.1 (2013): 105-119. *ulisboa.pt*. Web. 5. Apr. 2016. <jpl.letras.ulisboa.pt/articles/10.5334/jpl.79/galley/88/download/>
- Karttunen, Lauri, and Stanley Peters. "Conventional implicature." *Syntax and semantics 11: Presupposition*. Eds. Choon-Kyu Oh, and David A. Dinneen. New York:

- Academic Press, 1979. 1-56.
- Kauf, Carina. *Challenging Conditionals – A Semantical Analysis*. BA Thesis, Georg-August-Universität Göttingen, 2014.
- Klein, Wolfgang. *Time in language*. London: Routledge, 1994.
- Kocurek, Alexander W. *Counteridenticals*. Manuscript draft. 28 Oct. 2016. University of California, Berkely.
- Kratzer, Angelika. “Conditionals.” *Proceedings from the Regional Meeting of the Chicago Linguistic Society 2: Papers from the Parasession on Pragmatics and Grammatical Theory*. Eds. A. M. Farley, P. Farley, and K.-E. McCullough. Chicago: CLS, 1986. 1-15. *semanticsarchive.net*. Web. 30 July 2014. <<http://semanticsarchive.net/Archive/ThkMjYxN/Conditionals.pdf>>.
- , *Modals and Conditionals*. New York: Oxford UP, 2012.
- Kripke, Saul. “Identity and Necessity.” *Identity and Individuation*. Ed. M.K. Munitz. New York: New York UP, 1971. 135-164.
- , *Naming and Necessity*. Cambridge, MA: Harvard UP, 1980.
- Lakoff, George. “Sorry, I’m not myself today: The metaphor system for conceptualizing the self.” *Spaces, Worlds, and Grammar*. Eds. Gilles Fauconnier, and Eve Sweetser. Chicago: Chicago UP, 1996. 91-123.
- LaPorte, Joseph. “Rigid Designators.” *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition). Ed. Edward N. Zalta. Web. 21 Sep 2016. <<http://plato.stanford.edu/archives/spr2016/entries/rigid-designators/>>.
- Lewis, David. “Counterparts of persons and their bodies.” *The Journal of Philosophy* 68.7 (1971): 203-211.
- , *Counterfactuals*. Oxford: Blackwell, 1973.
- , “Attitudes De Dicto and De Se.” *Philosophical Review* 88 (1979): 513-543.
- Malamud, Sophia. *Semantics and pragmatics of arbitrariness*. Dissertation, University of Pennsylvania, 2006. Available at *ProQuest* (2006): AAI3225500.
- McCawley, James D. *Everything that Linguists Have Always Wanted to Know About Logic *but were ashamed to ask*. 2nd ed. Chicago: Chicago UP, 1993.
- Mikkelsen, Line. “Copular Clauses.” *Semantics: An International Handbook of Natural Language Meaning (HSK 33.2)*. Eds. Klaus von Heusinger, Claudia Maienborn, and Paul Portner. Berlin: de Gruyter, 2011. 1805-1829.
- Moltmann, Friederike. “Propositional attitudes without propositions.” *Synthese* 135.1 (2003): 77-118.
- Ninan, Dilip. *Imagination, Content and the Self*. PhD dissertation, MIT, 2008.
- Pearson, Hazel. *Counterfactual De Se*. Manuscript draft. December 2015. University of Tübingen/ZAS, Berlin.
- Pelletier, Jérôme. “Analogical Uses of the First Person Pronoun: a Difficulty in Philosophical Semantics.” *The Journal of Cognitive Science* 5.2 (2004): 139-155.

- Percus, Orin, and Uli Sauerland. “Pronoun movement in Dream Reports.” *Proceedings of NELS 33*. Eds. Makoto Kadowaki and Shigeto Kawahara. Amherst: GLSA UMass, 2003.
- Percus, Orin, and Yael Sharvit. “Copular asymmetries in belief reports.” Poster presented at *Semantics and Linguistics Theory (SALT) 24*. 2014. Available at <https://www.nyu.edu/projects/salt2014/SALT_2014/Program_files/percus-sharvit-salt2014-poster.pdf>
- Quadros, Ronice. Personal interview. 27 Apr. 2016.
- Quine, Willard van Orman. “Notes on existence and necessity.” *The journal of philosophy* 40.5 (1943): 113-127.
- , *Methods of logic*. 1950. New York: Henry Holt, 1959.
- , *Word and object*. Cambridge, MA: MIT Press, 1960.
- Recanatani, François. *Oratio Obliqua, Oratio Recta – An Essay on Metarepresentation*. Cambridge, MA: MIT P, 2000.
- Reichenbach, Hans. *Elements of symbolic logic*. New York: Macmillan, 1947.
- Stalnaker, Robert C. “A Theory of Conditionals.” *Studies in Logical Theory: American Philosophical Quarterly*. Ed. Nicholas Rescher. Oxford: Blackwell, 1968. 98-112.
- Stalnaker, Robert C., and Richmond H. Thomason. “A Semantic Analysis of Conditional Logic.” *Theoria* 36. 23-45.
- von Stechow, Kai. “The presupposition of subjunctive conditionals.” *The interpretive tract (MIT Working Papers in Linguistics 25)* Eds. Uli Sauerland and Orin Percus. Cambridge, MA: MITWPL, 1998. 29-44. *mit.edu*. Web. 1 Aug. 2016. <<http://mit.edu/fintel/fintel-1998-subjunctive.pdf>>.
- , “Subjunctive conditionals.” *The Routledge companion to philosophy of language*. Eds. Gillian Russell and Delia Graff Fara. New York: Routledge, 2012. 466-477.
- , Personal interview. 2 May 2016.
- Zalta, Edward N. “Gottlob Frege.” *The Stanford Encyclopedia of Philosophy* (Fall 2016 Edition). Ed. Edward N. Zalta. Web. 21. Sep 2016. <<http://plato.stanford.edu/archives/fall2016/entries/frege/>>.
- Zhang, Linmin. “Motivating a Symmetric Copula ‘Be’ in Attitude Reports.” *U Penn Working Papers in Linguistics* 22.1 Art. 34 (2016). Available at <<http://repository.upenn.edu/pwpl/vol22/iss1/34>>.
- Zobel, Sarah. *Impersonally Interpreted Personal Pronouns*. 2nd ed. Dissertation, Georg-August-Universität Göttingen, 2014.
- , Personal interview. 29 Sep. 2016.
- , *On the semantics of German als-phrases*. 2015. Manuscript in preparation.

Statutory Declaration

Ich versichere, dass ich die Arbeit selbständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen sind, sind als solche kenntlich gemacht. Ich versichere, dass die schriftliche (gebundene) und elektronische Form übereinstimmen.

Ort, Datum

Unterschrift