# The world is not enough: situations, laws and assignments in counterfactual donkey sentences

Dissertation submitted for the degree of
Doctor of Philosophy (Dr. Phil.)

presented by

*Andreas Walker*

at the

Universität Konstanz

Faculty of Humanities
Department of Linguistics

Konstanz, 2017

Tag der mündlichen Prüfung: 28. Juli 2017
1. Referentin: Prof. Dr. Maribel Romero
2. Referentin: JProf. Dr. María Biezma
3. Referentin: Prof. Dr. Ana Arregui

# Acknowledgements

A minimal situation of me writing this thesis contains very little: the thesis, me, and the writing relation. But if there is one thing to take away from this thesis, it is that our situations need to be a little richer. And indeed, there is a lot more that went into writing this thesis.

There is a number of people without whom this dissertation could not have been written. Maribel Romero, the best of all possible advisors, taught me how to think about formal semantics and how to teach others to think about it. Her greatest ability is probably to take her students seriously as young researchers, even if they do not do it themselves yet. Without her encouragement, curiosity and intellectual rigour, I doubt that I would have become a semanticist. Maribel convinced me to do a lot of things I did not want to do at first – from attending my first Amsterdam Colloquium to going to Santa Cruz (to say nothing about writing a thesis on counterfactual donkey sentences) – and she was always right. Her trust in my work kept me going when I doubted myself.

María Biezma, my (second) second advisor, always knew how to make me think more deeply about the things I was doing, looking beyond the lambdas and existential quantifiers. She insisted on the big picture and the story behind it all, and when I lost myself in the intricate details of an argument, she always reminded me that sometimes the common sense solution is actually right. If there is any coherence in this thesis, it is probably thanks to her.

Ana Arregui completed my dissertation committee as the external member, but her influence on this thesis was present long before that decision. Not only did her brilliant writing inspire many of my own ideas and provide a framework for my final proposal, but she also helpfully guided me through the philosophical minefields surrounding any engagement with situation semantics, spending much more time discussing my ideas with me than they probably deserve.

Beyond my committee, many people helped me shape my ideas by discussing my work with them. I can only highlight a few: Adrian Brasoveanu, who hosted my as a visitor at UCSC, and Pranav Anand, who re-taught me semantics when I was there. At Konstanz, either permanently or as visitors: Andrea Beltrama, Eva Csipak, Irene Heim, Sven Lauer, Brian Leahy and Doris Penka.

I cannot possibly list everyone else who taught me linguistics over the years, but there is a number of people I want to thank especially: Penka Stateva for getting me into semantics in the first place, Bettina Braun for always making me wonder whether phonetics would have been more interesting (and temporarily being my (first) second advisor), and Chiara Gianollo for showing me the beauty of historical linguistics.

I believe that one of the biggest dangers of writing a dissertation is to lose yourself in your research. There is a number of people who have taught me outside of linguistics and thus helped to keep me sane: Heike Drotbohm introduced me to the fascinating world of cultural anthropology and Uwe Jochum is responsible for me knowing way too much about theology, especially medieval ideas about the resurrection. Ralph Hafner introduced me to librarianship and Nico Kunkel dragged me back into the digital humanities debate. A lot of inspiration came from the fellow members of our research group *What if?*.

My sanity was also preserved by my fellow grad students. When I started out as the only semanticist, the phonetics and phonology crowd kindly took me in and listened to my talks about donkeys. Thank you Dani, Kathi and Jana especially. Later, finally more people saw the beauty

of semantics. Thank you, Arno, Antje, David, Erlinde, Felicitas, Gisela, Mark, Moritz, Ramona, Simone, Tina, Vasiliki and Yvonne – I'm going to miss you all. And I am also going to miss all my fellow Santa Cruz grad students, but especially Anna, Deniz and Kelsey.

My family deserves praise for bearing with all the interesting decisions I made in life, including studying linguistics. Special thanks to my sister Nathalie for being the most diligent proof-reader. I hope I can return the favor when she completes her dissertation, but I am afraid I am not as opinionated as her when it comes to the proper use of punctuation.

And of course none of this – really, none of this – would have been possible without the support and care from my friends and housemates over the years. They are too numerous to list, but I think they also know who they are. And most importantly: {Katharina, Karen}[1], a world without you is outside my epistemic possibilities.

[1] This is, as every good semanticist knows, an unordered set.

# Abstract (English)

In this thesis, I explore the phenomenon labeled **counterfactual donkey sentences** in the literature (van Rooij, 2006; Wang, 2009). Counterfactual donkeys combine the features of classical (indicative) donkey sentences (Geach, 1962) with features of counterfactual conditionals: They feature an indefinite noun phrase in the antecedent which is optionally anaphorically picked up by a pronoun in the consequent, as well as the morphological marking of counterfactual conditionals, e.g., an additional past tense in the antecedent and *would* in the consequent in English.

(1) If John had owned a donkey, he would have beaten it.

Counterfactual donkeys present a problem for similarity-based approaches to counterfactuals (Lewis, 1973b), because they sometimes require us to consider possibilities beyond the most similar ones, as the classical theory would have it: (1), for example, allows us to draw conclusions about John's attitude towards all donkeys, not just the ones he owns in the most similar worlds.

In this thesis, I compare three different solutions to this problem: First, I discuss a combination of Dynamic Predicate Logic (Groenendijk and Stokhof, 1991) with a variably strict semantics (Lewis, 1973b) and a dynamic strict semantics respectively (von Fintel, 1999), concluding that only the latter accounts for the full set of data. Second, I consider whether this solution can be extended to D-type theories (Elbourne, 2005) and conclude that this requires us to make stipulations that are not required in the dynamic framework. Third, prompted by additional data that shows the

same behaviour in the absence of an overt indefinite, I explore the option of accounting for counterfactual donkeys by switching to a situation-based framework that standardly considers a larger set of possibilities (Arregui, 2009), optionally restricting this set by law-like regularities in the context.

In evaluating all three proposals, I suggest that there are two possible further avenues: Exploring the empirical evidence for positing covert indefinites, allowing us to continue with a dynamic framework, or working out detailed accounts of pragmatic notions like relevance and law-like regularities, in order to precisify the situation-based framework.

# Abstract (German)

In dieser Dissertation befasse ich mich mit dem Problem, das in der bisherigen Literatur unter dem Begriff **counterfactual donkey sentences** verhandelt worden ist (van Rooij, 2006; Wang, 2009). Counterfactual donkeys kombinieren die Merkmale klassischer (indikativer) "donkey sentences" (Geach, 1962) mit Merkmalen kontrafaktischer Konditionale: Sie haben sowohl eine indefinite Nominalphrase im Antezedens, die optional anaphorisch von einem Pronomen im Konsequenz aufgegriffen werden kann, als auch die morphologische Markierung eines kontrafaktischen Konditionals, z.B. eine zusätzliche Vergangenheitsform im Antezedens und *would* im Konsequenz im Englischen.

(2) If John had owned a donkey, he would have beaten it.

Counterfactual donkeys werfen ein Problem für Ähnlichkeits-basierte Analysen von kontrafaktischen Konditionalen (Lewis, 1973b) auf, weil sie es manchmal nötig machen, Möglichkeiten in Betracht zu ziehen, die über die in der klassischen Theorie berücksichtigten ähnlichsten Möglichkeiten hinausgehen: (2) erlaubt es uns zum Beispiel, Schlüsse über die Einstellung Johns zu allen Eseln zu ziehen, nicht nur zu denjenigen, die er in den ähnlichsten Welten besitzt.

In dieser Dissertation vergleiche ich drei Lösungen für dieses Problem: Zuerst diskutiere ich eine Kombination von Dynamischer Prädikatenlogik (Groenendijk and Stokhof, 1991) mit jeweils einer variabel strikten Semantik (Lewis, 1973b) und einer dynamisch strikten Semantik (von Fintel, 1999) für kontrafaktische Konditionale, und komme zu dem Schluss, dass nur letztere alle Daten erklären kann. Zweitens betrachte ich, ob sich diese Lö-

sung auf D-Type-Theorien (Elbourne, 2005) erweitern lässt und komme zu dem Schluss, dass dies Stipulationen erfordert die in einer dynamischen Semantik nicht notwendig sind. Drittens motivieren zusätzliche Daten, die in der Abwesenheit eines sichtbaren indefiniten Artikels das gleiche Verhalten aufweisen, eine Analyse von counterfactual donkeys in einem situationsbasierten Framework das von Anfang an eine größere Menge an Möglichkeiten in die Evaluation von Konditionalen miteinbezieht (Arregui, 2009). Ich schlage vor, dass diese Menge optional durch gesetzesähnliche Regularitäten im Kontext eingeschränkt werden kann.

In der Evaluation aller drei Ansätze schlage ich vor, dass es zwei weitere Möglichkeiten gibt, die Forschung in diesem Bereich voranzubringen: Entweder eine Untersuchung der empirischen Evidenz für koverte Indefinite, die es uns erlaubt, das dynamische Framework zu verwenden, oder die Ausarbeitung einer detaillierten Analyse pragmatischer Begriffe wie Relevanz und gesetzesähnlicher Regularität, um das situationsbasierte Framework weiter zu präzisieren.

# Contents

# Chapter 1

# Introduction

## 1.1 Counterfactuals and similarity

In doing formal semantics, our starting point is relating sentences to their truth conditions. We take this to be a fairly good approximation to the notoriously difficult notion of *meaning*: Knowing the truth-conditions of a sentence might not guarantee that you fully understand its meaning, but understanding its meaning at least requires you to know its truth-conditions, that is, to be able to say under which circumstances a sentence is true or false. Consider (3):

(3) It is raining.

Imagine I am inside a room, with its windows shut and no sound coming in from the outside. You send me a text message from outside, saying (3). I can understand your message without knowing *whether* it is true or false; but I understand it by virtue of knowing how to establish its truth or falsity. I can do so by opening the windows and checking the actual state of the world. And I understand its meaning regardless of whether I actually perform this test, simply by virtue of knowing that this is the correct test to perforn. I understand the corresponding German sentence in (4) by virtue of the same method, showing me that it means the same as (3) – because it has the same truth conditions.

(4) Es regnet.

But now consider a sentence, a counterfactual conditional, like the following (5):

(5) If it hadn't rained, I would have stayed dry.

You enter my closed room, soaked to the bone, and utter (5). How am I supposed to establish its truth or falsity? What is the appropriate method? Intuitively, (5) suggests that it did in fact rain, and I can easily verify this by the same method I used to establish the truth and falsity of (3) and (4). But there is more to your sentence – clearly, it is not identical to (3) –, and this more is troubling. As it did in fact rain, how could I establish the truth or falsity of you staying dry *given the absence of rain*?

There is a standard answer to this question in the linguistic and philosophical literature, mostly known through the influential Stalnaker-Lewis theory (Lewis, 1973b; Stalnaker, 1968)[1]. The basic idea is to stick to our tried and tested method: You establish the truth and falsity of a sentence by comparing its truth conditions to the state of the world. But we will have to extend the scope of this method – where (3) and (4) require us to apply the method to the actual world (i.e., the world which you, I and our closed room inhabit), the sentence in (5) prompts us to look elsewhere. Specifically, we are invited to look at *other* possible worlds, which are very much like our own, but differ in some aspects. Clearly, they must differ in that they have to be worlds in which it didn't rain. But are they all such that you stay dry in them? This seems to depend on what further differences arise from this initial crucial difference. And the dynamics of such differences seem to be governed by which worlds – which collections of differences – we judge as more or less similar to our own world. Spelling out these dynamics, that is, giving a substantial theory of our similarity judgments – and our resulting truth value judgments for counterfactual conditionals like (5) – has been a major challenge for the Stalnaker-Lewis theory since its inception.

[1] A formal introduction to this theory will be given in Ch. 2.

## 1.2 Donkeys as a window into similarity

Giving a complete theory of similarity is an enterprise that extends far beyond the scope of this thesis. It presumably encompasses not only a semantics for counterfactuals, but also an account of our similarity judgments on the interface between pragmatics and psychology. Lacking this, we are always thrown back to Lewis's (1973) attitude towards similarity:

> Somehow, we *do* have a familiar notion of comparative overall similarity, even of comparative similarity of big, complicated, variegated things like whole people, whole cities or even – I think – whole possible worlds. However mysterious that notion may be, if we can analyze counterfactuals by means of it we will be left with one mystery in place of two. I am not one of those philosophers who seek to rest fixed distinctions upon a foundation quite incapable of supporting them. I rather seek to rest an unfixed distinction upon a swaying foundation, claiming that the two sway together rather than independently.
> — Lewis (1973b), p. 92.

A lot of Lewis's foundation will be left swaying in this thesis, although I do not share his pessimism in the long run – I am quite confident that our semantics will be supplemented with a well-founded theory of similarity judgments at some point. Until then, however, I think we can already fix a few points where the linguistic material alone at least seems to constrain our similarity judgments in a systematic manner, even while not entirely determining it. As I will argue, one such point is the interaction between indefinite noun phrases (and related constructions) in the antecedents of counterfactual conditionals with the similarity judgments we make in interpreting these sentences.

The data I will mostly be concerned with go back to observations by van Rooij (2006). They highlight a tension between the standard Lewis-Stalnaker semantics for counterfactuals and our intuitive judgments about sentences like (6):

(6) If John had owned a donkey, he would have beaten it.

The standard theory predicts that (6) is true if and only if the closest (most similar) worlds in which John owns a donkey are such that he beats that donkey in them. That is, the truth of (6) is established solely on the basis of those antecedent worlds which are most similar to ours. If some donkeys – let's say Platero and Grisella – are more likely to be owned by John, making worlds in which he owns them more similar to ours, then only his relationship with these two donkeys will figure in our evaluation of (6). So far the predictions.

As van Rooij (2006) points out, the most natural reading for (6), uttered out of the blue, is one in which we obtain the following entailments in (7):

(7) a. If John had owned donkey *a*, John would have beaten *a*.
b. If John had owned donkey *b*, John would have beaten *b*.
c. If John had owned donkey *c*, John would have beaten *c*.
d. If John had owned donkey *d*, John would have beaten *d*.
e. etc.

These entailments suggest a different picture: It seems that for any donkey in the domain of quantification, John's relationship with that donkey seems to matter. In Walker and Romero (2015) we call this reading the *high reading*. Examples of it can readily be found "in the wild", such as the following literary examples (all from the Gutenberg corpus of the Python Natural Language Toolkit, (Bird, Klein, and Loper, 2009)):

(8) If he were to enter into any profession with a view of better support, she would do all in her power to prevent him advancing in it. (Jane Austen, *Sense and Sensibility*)

(9) They believe that if a man picked a pocket he would naturally feel exquisitely good. (G.K. Chesterton, *The Man Who Was Thursday*)

Clearly, Austen's character is intending to prevent advancement not just in the most likely professions, but in all of them; and the belief described in Chesterton is one that extends to all men and pockets, not just those most likely to pick or be picked. But of course the usage of counterfactual donkey sentences with high readings is not limited to literary texts, but also occurs in more natural speech, as the following twitter data demonstrates[2]:

(10) if I wrote a joshler fic and posted it online where would you prefer to read it

(11) Maybe it would have been better if we met at a different period in our lives

(12) I'm convinced my life would be 100x better if I had a puppy to comfort me

(13) My luck is that bad lately. If I had a goldfish it would [d]rown.

Presumably, (10) is asking about a good place for all of their joshler fic, (11) is not referring solely to the most likely other period in their lives to meet, and we can conclude from (12) that if we offered them a puppy they would not reply with "I was not talking about that one". And (13)'s luck would turn out fatal for any goldfish they might acquire.

Note that this high reading co-exists with the more familiar reading in which these relationships don't matter. We call this second reading the *low reading* – but keep in mind that it simply corresponds to the standard theory reading – and observe that it is brought out by both richer contexts, as in (14) and certain structures, like the identificational sentences in (15) (from van Rooij (2006)) and (16) (from Twitter):

(14) John hates all donkeys except for his grandfather's donkey Melissa, which he loves. On Tuesday, John was supposed to inherit Melissa but declined the will. But if he had inherited a donkey on Tuesday, he would have loved it.

[2] The data cited below was gathered via the Twitter API through a Python script. The following sentences are excerpts, without the user names and/or additional material.

(15) If Alex had been married to a girl from his class, it would have been Sue.

(16) but if we were talking about a heartache... it would be you...

Note that while this phenomenon was first discussed in the context of counterfactual donkey sentences, it actually extends beyond conditionals with the classical donkey structure, that is, specifically, beyond conditionals with an indefinite in the antecedent and a pronoun in the consequent. First, no pronoun is required (and indeed, none of the literature discussed in this thesis derives any facts about high and low readings from the presence of a pronoun):

(17) If John had owned a donkey, Mary would have been happy.

The sentence in (17) can be taken to either mean that for any donkey, if John had owned that donkey, Mary would have been happy (the high reading). Or, in a context like (14), it could mean that Mary would have been happy, because John would most likely have inherited a particular donkey (the low reading).

But not even the indefinite is obligatory, as the following example, due to Schwarz (2013), shows:

(18) If the dart had landed on the left side of the board, I would have won.

Again, we yield a high reading (in which for any place on the left side of the board, the dart landing on it would have been sufficient for me winning), although there is no indefinite in the antecedent to trigger it.

How can we explain these two readings, whatever is responsible for them arising? One way would be to stipulate that in the case of the high reading, the similarity ordering in question simply does not have the shape described above. Another way is to say that our picture of similarity as layed out in the standard theory is incomplete, and that we have found a fixed point that tells us something about the constraints that are at work in our similarity judgments. In explaining how the entailments in (7) arise, we will also be able to explain part of how similarity works.

It should be obvious that in what follows, I am going to pursue the latter. However, I believe that the first option is one that needs to be taken seriously, and which needs to – and can – be argued against empirically. I take Wang (2009) to implicitly take the position I am going to sketch in the following, but for fairness' sake I will have to admit that none of the views I attribute to her are explicitly spelled out in her paper. Rather, they are my own extrapolations of what I believe to be necessary if her view is to explain the empirical observations.

## 1.3 Similarity should be principled

Wang (2009) is a reply to van Rooij's (2006) account which we are going to inspect in detail in Ch. 2. In this reply, Wang upholds that the entailments arising from (6) are not a distinct (high) reading that requires additional machinery for its explanation, but rather that they are well accounted for in the standard theory[3]. Unfortunately, she then proceeds to discuss a number of cases that all lack the entailments observed by van Rooij (2006) – that is, do not seem to have a high reading to begin with – and never returns to cases like (6). What, then, would it take to account for the entailments in (7) within Wang's account?

Here is what I think it would take: In order for the entailments to arise, we need to stipulate that our similarity relation is of the shape described in (19).

(19) Wang's Stipulation The set of closest worlds contains at least one world for each referent of the indefinite in which this referent satisfies the antecedent.

[3] Wang (2009) spells out her analysis in an update semantics in the tradition of Veltman (2005). For simplicity, we drop this aspect in the following discussion, as it does not seem to bear on the issue at hand.

This is a tough stipulation to swallow in the absence of independent motivation. How could we justify such a stipulation? And why should this stipulation only hold in certain cases and not others? Clearly, this gives us the right empirical predictions – as far as it predicts anything –, but it does nothing to explain them. But we could be tempted to reformulate our enterprise as follows: An explanation of the entailments in (7) consists in an explanation of Wang's stipulation. Then, again, we could put away van Rooij's (2006) machinery and remain within the standard theory. But this is not the right way either: As it turns out, there are cases in which, arguably, Wang's stipulation explicitly does *not* hold, and yet we observe the same entailments as in (6)[4].

In order to show that Wang's stipulation is an unlikely source of these entailments, we will proceed as follows: First, we will set up a scenario with explicit assumptions about the probability of certain referents for the indefinite satisfying the antecedent. Note that I am using probability in an entirely non-theoretical sense here, and will not engage with probabilistic semantics at all. But as it turns out, talk about probability – in this lay sense – comes very close to talk about similarity, at least in certain contexts, and so we will use it as a way of shaping our scenarios in the required way[5]. Second, having set up that scenario, we will show that it nonetheless allows for high readings. This requires some care in constructing our target sentences, as we are faced with the unfortunate fact that high readings usually entail low readings, making a truth value judgment undecisive in these cases. However, we can use both negation and the existential quantification of *might* to get around this problem.

(20) Scenario: There are two farmers in the kingdom of King Kakos, called Onophilos and Onophobos. Both are very poor and do not own a donkey. Onophobos is a cruel man who would love to own and beat a donkey. He has been saving money all his life and has

[4] Maribel Romero and I previously published this argument in Walker and Romero (2015, 2016).

[5] This decision may not be entirely uncontroversial, but strikes me as a good heuristics until definitely proven wrong, both in virtue of its intuitive plausibility and its usefulness in accounts like Lewis (2016).

nearly enough to buy a donkey. Onophilos is a mild-mannered vegan who has no means to acquire or interest in owning a donkey, much less so in beating it. King Kakos only knows Onophobos and is convinced that all inhabitants of his kingdom are just as cruel and evil as Onophobos. He discusses this with his advisor, who is well-informed about all the farmers in the kingdom and their dispositions.

(21) King Kakos: Here's what I think about the farmers in my kingdom. If a farmer in a kingdom owned a donkey, he would beat it.

(22) Advisor: You are wrong. It's not the case that if a farmer in your kingdom owned a donkey, he would beat it. Onophilos, for example, is a vegan and would never do so.

In the context of (20), we judge the advisor's claim in (22) as true. However, we have explicitly set up a context in which a world where Onophilos owns a donkey is much less similar than one in which Onophobos does. The prediction of Wang's approach is then that either (22) should be judged as false, or that the technical notion of similarity judgments employed in counterfactuals deviates so much from our intuitive similarity judgments that it becomes nearly unusable. Note that the advisor's claim is not simply a low reading that somehow has made Onophilos the more likely recipient of a donkey either: Clearly, he can felicitously follow up on (22) with the sentence in (23), showing that he is generalizing about all farmers[6].

(23) But neither is it the case that if a farmer in your kingdom owned a donkey, he would treat it well. Onophobos would beat his donkey relentlessly.

The example in (22) is somewhat convoluted, as it relies on embedding the counterfactual statement under negation in order to tease apart the high and low readings. In Walker and Romero (2016), we consider two more environments in which it is possible to tell them apart: One in which the (default universal) quantificational force in the conditional is made

[6] Thanks to Amy Rose Deal (p.c.) for pointing out this potential counterargument.

overt and negated, and one in which a *might*-conditional is used instead. I present both here, although somewhat tentatively: At least the latter relies on treating *might* as the dual of *would*, and I want to remain agnostic with respect to this point.

(24) If a farmer in your kingdom owned a donkey, he wouldn't necessarily beat it. Onophilos, for example, is a vegan and would never do so.

(25) If a farmer in your kingdom owned a donkey, he might not beat it. Onophilos, for example, is a vegan and would never do so.

As this brief detour through Wang's (2009) challenge demonstrates, if we want to uphold a principled account of similarity as the basis for our judgments of counterfactuals' truth, there is something more to the high reading than a mere stipulation. We do not want similarity to be so unconstrained that it can generate just about any reading; I furthermore would like similarity to not deviate too much from what we would intuitively call a similarity judgment. This does not have to be a *naive* judgment – as Lewis (1979) points out, e.g., we may have to give different weights to facts and laws –, but it should also not be a relation that is entirely inaccessible to the observer, hand-tailored to produce the desired results.

Further evidence for the principledness of similarity comes from the distribution of high and low readings. As van Rooij (2006) points out, the high reading is strongly preferred in the absence of context; however, there are many contexts that very easily push us towards preferring a low reading. And finally, in constructions like (15), the identificational cases, we seem to be faced with an *obligatory* low reading. The following table provides an overview, where # is dispreferredness and * is unavailability:

| | out of the blue | low context | identificational sentence |
|---|---|---|---|
| high | ✓ | # | * |
| low | # | ✓ | ✓ |

In explaining high and low readings, we should also attempt to link their mechanics to features of the context that govern this distribution. In the end, we will emerge with a picture where similarity is highly context-dependent – but in a principled way, rather than by stipulation or the magic of similarity relations selected *ad hoc*. Counterfactual donkey sentences – the heading under which this data has been discussed in the literature so far – offer us a window into the workings of similarity, in a way that will generalize beyond them.

## 1.4 Outline of the thesis

Counterfactual donkey sentences cannot be analyzed without taking into account the theory of donkey sentences in more general terms. There are two major lines currently under consideration in the literature, dynamic semantics (represented by, e.g., Kamp (1981), Heim (1982), Groenendijk and Stokhof (1991), Brasoveanu (2013), a.o.) and D-type theory (represented by, e.g., Büring (2004), Elbourne (2005, 2013), a.o.).

The literature on counterfactual donkeys so far (van Rooij, 2006; Walker and Romero, 2015, 2016; Wang, 2009) has looked at them through the lense of dynamic semantics, building on van Rooij's (2006) original proposal for combining Groenendijk and Stokhof's (1991) Dynamic Predicate Logic (DPL) with Lewis-Stalnaker semantics. I will introduce and evaluate this approach in Ch. 2 in the following way: After presenting Groenendijk and Stokhof's (1991) classical results on donkey sentences in DPL and giving a motivated formal introduction to Lewis's (1973) variably strict semantics for counterfactuals, I will then proceed to show how van Rooij (2006) combines them in order to yield the high reading and extend on his presentation by providing both an accessible way of visualizing his results and a formal demonstration of how his proposal plays out in a DPL running on world-assignment pairs. I will then introduce Negative Polarity Items (NPI) and show how they present a challenge to van Rooij's proposal. To overcome this challenge, I will present von Fintel's (1999)

dynamic strict semantics and show how it can be combined with DPL, preserving van Rooij's results while also licensing NPIs correctly. This last section builds on results from Walker and Romero (2015) and refines them. In the discussion of this section, I argue that this method makes correct empirical predictions, although at a cost: It both commits us to a dynamic analysis of donkey sentences and to the stipulation of covert indefinites to account for high readings in cases where no overt indefinite is available.

I will then address both of these costs in turn. In Ch. 3, I explore the possibility of transposing the mechanisms from the dynamic solution to Elbourne's (2005) D-type theory. For this purpose, I will first introduce classical D-type theory. I then continue to show that importing dynamic insights into this framework is not trivial: First, addressing the problem of proportional and weak readings (Kadmon, 1987; Schubert and Pelletier, 1987) requires us to either accept empirical shortcomings, make radical (and potentially problematic) changes to our semantics for generalized quantifiers or import so much dynamic machinery that our D-type theory will end up indistinguishable from dynamic proposals. Second, in extending D-type theory to counterfactuals, we need to stipulate mechanisms that are less clearly motivated than the use of assignments in dynamic semantics. In the discussion, I conclude that a proposal that wants to derive the facts about high and low readings from the mechanisms used to explain indicative donkey sentences has better chances of succeeding if it employs dynamic semantics.

In Ch. 4, I then turn to the second concern: Do we want to use the machinery of indicative donkey sentences to account for high and low readings to begin with? This strategy crucially relies on the presence of indefinites in both types of sentences. However, as the empirical data shows, we also obtain high and low readings in sentences that do not display an overt indefinite or indefinite-like structure (such as disjunction). For this reason, I explore an alternative strategy here: Rather than using the resources of donkey sentences, I turn to the resources provided to us by the literature on counterfactuals. I begin by discussing three types of variations on the Lewis-Stalnaker semantics: Lewis's (2016) and Nichols's

(2016) relevance-based accounts, Schulz's (2011) causal account and Arregui's (2009) local similarity account. I then consider the problem of high and low readings from the perspective of Arregui's account: Starting from a larger domain of quantification to begin with standardly yields the high reading. I argue that those contexts that facilitate low readings do so by introducing regularities that can be elevated to law-like status in evaluating a counterfactual. Such regularities can then constrain the space of quantification, yielding a low reading. I conclude that this allows us to account for the empirical data without relying on the presence of an indefinite. However, this framework is so radically context-dependent that we might run the risk of overgeneration in the absence of a constraining pragmatic theory.

In Ch. 5, I conclude the thesis with an evaluation of both available strategies, proposing to further investigate two issues: First, the problem of covert indefinites might be alleviated if their presence could be shown independently. Second, the potential overgeneration issues could be addressed by formulating a more constraining pragmatic theory that takes into account, e.g., facts about the behaviour of counterfactual donkey sentences in larger discourses.

# Chapter 2

# The dynamic line

In this chapter, I will explore a potential solution to the problem of high and low readings in the framework of dynamic semantics, building on the original proposal by van Rooij (2006). Van Rooij's proposal is a combination of two ingredients: Groenendijk and Stokhof's (1991) Dynamic Predicate Logic (DPL) and Lewis's (1973) variably strict conditional analysis of counterfactuals. We will begin with a brief sketch of both accounts and their motivations. After showing how van Rooij (2006) derives high readings from his analysis – the de facto standard solution in the literature up to now –, I will then present a problem for his account: While it is one of van Rooij's (2006) goals to account for the licensing of Negative Polarity Items (NPI) in the antecedents of counterfactuals, he actually only succeeds in doing so for high readings. However, as I will demonstrate, NPIs are also licensed in the antecedents of low counterfactual donkey sentences. In order to solve this problem, I propose reformulating van Rooij's insights in a dynamic strict semantics (von Fintel, 1999, 2001), refining a solution previously sketched in Walker and Romero (2015). I will show that this successfully generates high and low readings, as in van Rooij's (2006) account, while inheriting successful NPI licensing in all relevant instances from von Fintel's account. In the discussion, however, I point out that

this is not enough: While we get the two readings right, we still lack a systematic account of their distribution. In following van Rooij's solution, we also commit ourselves to a dynamic semantics, which may conflict with other theoretical desiderata.

## 2.1 Dynamic Predicate Logic

### 2.1.1 The problem of donkey sentences

Dynamic Predicate Logic (Groenendijk and Stokhof, 1991) is a compositional variant of the dynamic semantics originally proposed by Heim (1982) and Kamp (1981). Its main purpose is addressing the puzzle of donkey sentences, like (26), and the related puzzle of intersentential anaphora, as in (27)[1].

(26) If $a^x$ farmer owns $a^y$ donkey, $he_x$ beats $it_y$.

(27) $A^x$ man walked in. $He_x$ was wearing $a^y$ hat.

Sentences like (26) are puzzling from the perspective of a traditional static semantics because of the indefinite's behaviour in them. Outside of conditional sentences, we make the standard assumption that indefinite articles correspond to existential quantification, as in (28) and its translation in (29):

(28) $A^x$ farmer owns $a^y$ donkey.

(29) $\exists x, y : farmer(x) \wedge donkey(y) \wedge own(x, y)$

However, applying this to the conditional sentence in (26) gives us the incorrect truth conditions in (30):

(30) $(\exists x, y : farmer(x) \wedge donkey(y) \wedge own(x, y)) \rightarrow beat(x, y)$

[1] In this chapter, I'll be following the convention of dynamic semantics to indicate the variables introduced or anaphorically picked up by an expression with super- and subscripts respectively

There are two related problems with (30), one formal and one in terms of interpretation: Formally, the variables $x$ and $y$ in the consequent of the conditional remain unbound, as they are outside the scope of the existential quantifier. Yet even if we somehow managed to bring them into the scope of the quantifier, we would still end up with the wrong truth conditions, as (31) demonstrates:

(31) $\exists x, y : ((farmer(x) \wedge donkey(y) \wedge own(x, y)) \rightarrow beat(x, y))$

The formula in (31) is true as soon as we find a pair of individuals that is not a farmer and a donkey standing in the owning-relationship, as the conditional in propositional logic is true as soon as its antecedent is false. This will trivially be satisfied in most models, by virtue of discovering any two individuals that do not make the conjunction in (31) true. The correct truth conditions for (26) instead involve universal quantification:

(32) $\forall x, y : ((farmer(x) \wedge donkey(y) \wedge own(x, y)) \rightarrow beat(x, y))$

That is, the sentence in (26) asserts that for all pairs of farmers and donkeys that stand in the owning-relationship, the farmer beats the donkey. The challenge for a theory of donkey sentences is to derive the intuitively correct truth conditions in (32) compositionally from the linguistic material present in (26) without giving up on the standard interpretation of the indefinite in the non-conditional case in (28).

The proposed solution in Dynamic Predicate Logic is to leave the standard compositional process unchanged – that is, to compositionally derive the, at first sight incorrect, truth conditions in (30) –, but to adapt the underlying semantics of predicate logic in such a way that (30) ends up being equivalent to (32):

(33) $\exists x Px \rightarrow Qx \equiv \forall x(Px \rightarrow Qx)$

### 2.1.2 The dynamic turn

Crucially, DPL moves from a picture where the interpretation of logical formulas is dependent on contextual assignments to a picture where the interpretation of logical formulas also influences these contextual assignments in turn. In a static semantics, we can think of the meaning of a sentence as the set of contextual assignments that verify it, as in (34): As the interpretation of $\exists xPx$, we obtain the set of assignments $g$ for which there exists an assignment $h$ that differs from $g$ at most in the value of $x$ such that the individual assigned to $x$ in $h$ is contained in the interpretation of $P$[2] .

(34) $[\![\exists xPx]\!] = \{g \mid \exists h : h[x]g \wedge h(x) \in F(P)\}$

However, as is obvious from (34), the changes made by existential quantification – changing the value of $x$ to obtain the updated assignment $h$ – are not reflected in the set obtained from the formula. This is where dynamic semantics innovates: It treats sentences as pairs of assignments. Each pair consists of the original input assignment $g$ and the resulting output assignment $h$. This way, updating the assignment function can have a lasting effect on the context by passing on the output assignment of a formula as the input assignment for the next formula. For existential quantification, in parallel to (34), we obtain the following interpretation:

(35) $[\![\exists xPx]\!] = \{\langle g, h\rangle \mid h[x]g \wedge h(x) \in F(P)\}$

In (35), existential quantification has the same effect as in the static formula above. However, the updated assignment $h$ is explicitely recorded as the output assignment. The DPL solution to the puzzle of donkey sentences is based on the availability of these output assignments: By passing them on within the conditional, we can derive the equivalence in (33). The semantics for predicate logic are defined as follows in Groenendijk and Stokhof (1991):

[2] Like Groenendijk and Stokhof (1991), I assume an interpretation function $F$ that maps individuals to individual constants and sets of $n$-tuples of individuals to $n$-place predicates.

(36) $[\![Rt_1...t_n]\!] = \{\langle g, h\rangle \mid h = g \wedge \langle [\![t_1]\!]^h ... [\![t_n]\!]^h\rangle \in F(R)\}$

(37) $[\![\exists x\phi]\!] = \{\langle g, h\rangle \mid \exists k : k[x]g \wedge \langle k, h\rangle \in [\![\phi]\!]\}$

(38) $[\![\phi \wedge \psi]\!] = \{\langle g, h\rangle \mid \exists k : \langle g, k\rangle \in [\![\phi]\!] \wedge \langle k, h\rangle \in [\![\psi]\!]\}$

(39) $[\![\phi \rightarrow \psi]\!] = \{\langle g, h\rangle \mid h = g \wedge \forall k : \langle h, k\rangle \in [\![\phi]\!] \rightarrow \exists j : \langle k, j\rangle \in [\![\psi]\!]\}$

(40) $[\![\forall x\phi]\!] = \{\langle g, h\rangle \mid h = g \wedge \forall k : k[x]g \rightarrow \exists j : \langle k, j\rangle \in [\![\phi]\!]\}$

With the ingredients in (36) - (40), we can now show that the postulated equivalence in (33) holds, assigning the correct truth conditions to donkey sentences:

(41) a. $[\![\exists xPx]\!]$

$= \{\langle g, h\rangle \mid \exists k : k[x]g \wedge \langle k, h\rangle \in [\![Px]\!]\}$ by (37)

$= \{\langle g, h\rangle \mid \exists k : k[x]g \wedge k = h \wedge [\![x]\!]^h \in F(P)\}$ by (36)

$= \{\langle g, h\rangle \mid h[x]g \wedge h(x) \in F(P)\}$ by $k = h$

b. $[\![\exists xPx \rightarrow Qx]\!]$

$= \{\langle g, h\rangle \mid h = g \wedge \forall k : \langle h, k\rangle \in [\![\exists xPx]\!]$
$\rightarrow \exists j : \langle k, j\rangle \in [\![Qx]\!]\}$ by (39)

$= \{\langle g, h\rangle \mid h = g \wedge \forall k : (k[x]h \wedge k(x) \in F(P))$
$\rightarrow \exists j : \langle k, j\rangle \in [\![Qx]\!]\}$ by (41a)

$= \{\langle g, h\rangle \mid h = g \wedge \forall k : (k[x]h \wedge k(x) \in F(P))$
$\rightarrow \exists j : k = j \wedge j(x) \in F(Q)\}$ by (36)

$= \{\langle g, h\rangle \mid h = g \wedge \forall k : (k[x]h \wedge k(x) \in F(P))$
$\rightarrow k(x) \in F(Q)\}$ by $k = j$

c. $[\![Px \rightarrow Qx]\!]$
$= \{\langle g, h\rangle \mid h = g \wedge \forall k : \langle h, k\rangle \in [\![\phi]\!]$
$\rightarrow \exists j : \langle k, j\rangle \in [\![Qx]\!]\}$ by (39)

$= \{\langle g, h\rangle \mid h = g \wedge \forall k : (h = k \wedge k(x) \in F(P))$
$\rightarrow \exists j : (k = j \wedge j(x) \in F(Q))\}$ by (36)

$= \{\langle g, h\rangle \mid h = g \wedge ((h(x) \in F(P))$
$\rightarrow (h(x) \in F(Q)))\}$ by $k = j$ and $h = k$

d. $[\![\forall x(Px \rightarrow Qx)]\!]$
$= \{\langle g, h\rangle \mid h = g \wedge \forall k : k[x]g$
$\rightarrow \exists j : \langle k, j\rangle \in [\![Px \rightarrow Qx]\!]\}$ by (40)

$= \{\langle g, h\rangle \mid h = g \wedge \forall k : k[x]g$
$\rightarrow \exists j : k = j \wedge (j(x) \in F(P)) \rightarrow (j(x) \in F(Q))\}$ by (41c)

$= \{\langle g, h\rangle \mid h = g \wedge \forall k : k[x]g$
$\rightarrow (k(x) \in F(P)) \rightarrow (k(x) \in F(Q))\}$ by $k = j$

$= \{\langle g, h\rangle \mid h = g \wedge \forall k : (k[x]g \wedge k(x) \in F(P))$
$\rightarrow k(x) \in F(Q)\}$ by Exportation

$= \{\langle g, h\rangle \mid h = g \wedge \forall k : (k[x]h \wedge k(x) \in F(P))$
$\rightarrow k(x) \in F(Q)\}$ by $h = g$

Compare the last lines of (41b) and (41d) to see that they are indeed equivalent.

Note that this semantics only models the dynamic behaviour of the anaphoric potential of linguistic expressions. It is not a full discourse model and does not allow us to model the development of the common ground over time. We will return to this issue in 2.2.3 below.

## 2.2 Variably strict counterfactuals

### 2.2.1 The Stalnaker-Lewis analysis

The proposal due to van Rooij (2006) starts from a straightforward combination of DPL with the Stalnaker-Lewis analysis of counterfactuals. Since we have only sketched this analysis informally in Ch. 1, we will first give a rigorous definition of the relevant concepts, based on Lewis's (1973) version.

Regardless of its status as an appropriate analysis for indicative conditionals – a debate which lies far outside the scope of this dissertation –, it is very clear that the material conditional familiar from propositional logic is not an appropriate analysis for counterfactual conditionals: As their antecedents are false by definition, we would be predicting that all counterfactual conditionals are judged true (see (42)). This is obviously not the case.

(42)

| $p$ | $q$ | $p \rightarrow q$ |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 0 | 0 |
| 0 | 1 | 1 |
| 0 | 0 | 1 |

The standard answer to this problem is to make the truth of the counterfactual depend on the truth of the material conditional in worlds different from the actual world. That is, the counterfactual can be understood, tentatively, as a material conditional embedded under a modal necessity operator, as in (43). This is equivalent to (44), with $D_w$ denoting the domain of worlds accessible from $w$. This analysis of the conditional is known as the strict conditional analysis.

(43) $[\![\textit{if } p \textit{ would } q]\!]^w = 1$ iff $\Box(p \rightarrow q)$

(44) $[\![\textit{if } p \textit{ would } q]\!]^w = 1$ iff $\forall w' \in D_w : w' \in [\![p]\!] \rightarrow w' \in [\![q]\!]$

However, as Lewis (1973b) points out, there is no accessibility relation that will generate the correct truth conditions for counterfactuals, as the worlds to be taken into consideration crucially depend on the content of the antecedent. He demonstrates the problem with what has become known as *Sobel sequences*:

(45) a. If the USA threw its weapons into the sea tomorrow, there would be war.

b. But if the USA and the other nuclear powers all threw their weapons into the sea tomorrow, there would be peace.

c. But if they did so without sufficient precautions against polluting the world's fisheries, there would be war.

d. But if, after doing so, they immediately offered generous reparations for the pollution, there would be peace.

The formal structure corresponding to the – intuitively felicitous – sequence in (45) is the following:

(46) a. $\phi_1 \boxrightarrow \psi$

b. $\phi_1 \wedge \phi_2 \boxrightarrow \neg\psi$

c. $\phi_1 \wedge \phi_2 \wedge \phi_3 \boxrightarrow \psi$

d. $\phi_1 \wedge \phi_2 \wedge \phi_3 \wedge \phi_4 \boxrightarrow \neg\psi$

Applying any strict analysis of the conditional to (46a), we obtain the following:

(47) $\forall w' \in D_w : w' \in [\![\phi_1]\!] \rightarrow w' \in [\![\psi]\!]$

Whatever accessibility relation – that is, domain $D_w$ – we choose, (47) asserts that all accessible worlds that verify $\phi_1$ also verify $\psi$. However, the accessible worlds that verify $\phi_1 \wedge \phi_2$ – the antecedent of (46b) – are a subset of those worlds. Accordingly, they must also all satisfy $\psi$, contradicting (46b). This problem arises for any two adjacent counterfactual conditionals in a Sobel sequence. It is obvious that the accessibility relation cannot be held constant between counterfactuals. As a solution, Lewis (1973b) proposes his variably strict semantics:

(48) $[\![\textit{if } p \textit{ would } q]\!]^w = 1$ iff $\forall w' \in f_w([\![p]\!]) : w' \in [\![q]\!]$

In contrast to (44), (48) replaces the fixed domain $D_w$ with a selection function that returns worlds based on the antecedent $p$:

(49) $f_w([\![p]\!]) = \{v \in [\![p]\!] \mid \neg\exists u \in [\![p]\!] : u <_w v\}$

That is, the domain now consists of those worlds $v$ which satisfy the antecedent $p$ and for which there is no other antecedent world $u$ that is closer to the actual world than $v$, where closeness is defined over the contextually given similarity ordering $<_w$. Since any counterfactual in a Sobel sequence will have its own domain of accessible worlds – based on the content of its antecedent –, we avoid this problem of the strict conditional approach.

The similarity ordering $<_w$ ranks worlds by their similarity to the actual world. It is reflexive, transitive and connected, yielding a spherical setup of worlds: The actual world is maximally similar to itself and surrounded by spheres of possible worlds, with all worlds within a sphere equally similar to the actual world, and similarity decreasing from sphere to sphere. (50) provides a linearized representation of this spherical model, with similarity decreasing from left to right (that is, from table cell to table cell, but with worlds within the same cell being equally similar to $w_0$.) The selection function returns only and all of the antecedent-worlds from the first sphere that contains an antecedent-world.

(50)

| | | | | |
|---|---|---|---|---|
| $w_0$ | $w_1$ | $w_2$, $w_3$, $w_4$ | $w_5$, $w_6$ | $w_7$, $w_8$ |

Now assume that $p = \{w_2, w_3, w_5, w_7, w_8\}$, indicated by boldface in (51):

(51)

| | | | | |
|---|---|---|---|---|
| $w_0$ | $w_1$ | $\mathbf{w_2}$, $\mathbf{w_3}$, $w_4$ | $\mathbf{w_5}$, $w_6$ | $\mathbf{w_7}$, $\mathbf{w_8}$ |

Then the selection function for $p$, $f_w([\![p]\!])$, selects the first sphere – from left to right – that contains an antecedent world, in this case the third sphere. It returns all antecedent-worlds from that sphere, that is, $w_2$ and $w_3$ in our example:

(52)

| | | | | |
|---|---|---|---|---|
| $w_0$ | $w_1$ | $\mathbf{w_2}$, $\mathbf{w_3}$, $w_4$ | $\mathbf{w_5}$, $w_6$ | $\mathbf{w_7}$, $\mathbf{w_8}$ |

The antecedent-worlds in the fourth and fifth sphere are ignored, as is the non-antecedent world $w_4$ within the third sphere.

### 2.2.2 Two-dimensional similarity

As discussed in Ch. 1, Lewis's (1973) variably strict approach only predicts what we call the low reading. This follows straightforwardly from its semantics, as defined in (48) and (49): Since we are only concerned with the validity of the material conditional in the most similar antecedent-worlds, if there is a particularly likely way of verifying the antecedent, we will only be concerned with the validity of the material conditional in worlds where the antecedent is satisfied in that particular way. However, the entailments in (7) for (6) – repeated below as (53) and (54) – suggest that we do in fact need to consider at least one world per individual that satisfies the indefinite noun phrase in the high reading.

(53) If John had owned a$^x$ donkey, he$_j$ would have beaten it$_x$.

(54) a. If John had owned donkey *a*, John would have beaten *a*.
b. If John had owned donkey *b*, John would have beaten *b*.
c. If John had owned donkey *c*, John would have beaten *c*.
d. If John had owned donkey *d*, John would have beaten *d*.
e. etc.

As van Rooij (2006) points out, this suggests an equivalence very similar to the one observed in dynamic semantics for indicative donkey sentences – see (33) –, but in the counterfactual domain:

(55) $\exists x Px \mathrel{\Box\!\!\rightarrow} Qx \equiv \forall x(Px \mathrel{\Box\!\!\rightarrow} Qx)$

In order to obtain the high reading, van Rooij (2006) combines Lewis's (1973) variably strict semantics with DPL. Where in the standard static semantics – and consequently, the Lewisian analysis –, the meaning of a sentence could be thought of as a set of worlds (i.e. the worlds that verify the truth conditions), the dynamic analysis treats sentences as functions

from sets of (input) assignments to sets of (output) assignments. Combining these views yields a picture in which sentence meanings are functions from world-assignment pairs to world-assignment pairs, where the update can both be eliminatory (removing world-assignment pairs that do not verify the sentence) and enriching (where the sentence updates the assignment, e.g. in the case of existential quantification).

Consequently, in considering the semantics of counterfactuals, we need to reconsider similarity: Instead of simply relating worlds, it now has to relate world-assignment pairs. And instead of the one-dimensional ordering in (50), we now obtain a set of orderings that can be represented by a two-dimensional picture, with the original sphere model on the x-, and the set of input assignments on the y-axis. Note that the assignments in this example essentially correspond to the various donkeys the indefinite could refer to, that is, for (53), we can assume that $g_1 = g^{a/x}, g_2 = g^{b/x}$, etc. As worlds are ordered by similarity, and as we inherit the original ordering, the x-axis in the picture is ordered by similarity from left to right, essentially a copy of the one-dimensional picture in (50). However, assignments are not ordered in any particular way, so that the order of the elements on the y-axis is arbitrary.

(56)

| | | | | | |
|---|---|---|---|---|---|
| $g_4$ | | | | | |
| $g_3$ | | | | | |
| $g_2$ | | | | | |
| $g_1$ | | | | | |
| | $w_0$ | $w_1$ | $w_2$, $w_3$, $w_4$ | $w_5$, $w_6$ | $w_7$, $w_8$ |

In a system like the one represented by (56), the antecedent of a counterfactual is not verified by a world alone, but by a world-assignment pair. Assuming that $p$ is verified, for example, by $\{\langle w_2, g_1\rangle, \langle w_3, g_2\rangle, \langle w_5, g_3\rangle, \langle w_5, g_2\rangle, \langle w_7, g_4\rangle, \langle w_8, g_4\rangle\}$, we can represent the domain for our selection function as follows:

(57)

| | | | | | |
|---|---|---|---|---|---|
| $g_4$ | | | | | $\langle w_7, g_4\rangle$, $\langle w_8, g_4\rangle$ |
| $g_3$ | | | | $\langle w_5, g_3\rangle$ | |
| $g_2$ | | | $\langle w_3, g_2\rangle$ | $\langle w_5, g_2\rangle$ | |
| $g_1$ | | | $\langle w_2, g_1\rangle$ | | |
| | $w_0$ | $w_1$ | $w_2$, $w_3$, $w_4$ | $w_5$, $w_6$ | $w_7$, $w_8$ |

But now, in contrast to the one-dimensional picture, we are faced with a choice. Our selection function can either remain a classical Lewisian one, returning only the world-assignment pairs from the first sphere (now represented by an entire column instead of a single cell), as in (58). This yields the standard low reading.

(58)

| | | | | | |
|---|---|---|---|---|---|
| $g_4$ | | | | | $\langle w_7, g_4\rangle$, $\langle w_8, g_4\rangle$ |
| $g_3$ | | | | $\langle w_5, g_3\rangle$ | |
| $g_2$ | | | $\langle w_3, g_2\rangle$ | $\langle w_5, g_2\rangle$ | |
| $g_1$ | | | $\langle w_2, g_1\rangle$ | | |
| | $w_0$ | $w_1$ | $w_2$, $w_3$, $w_4$ | $w_5$, $w_6$ | $w_7$, $w_8$ |

Alternatively, we can let the selection function select the world-assignment pairs from the leftmost cell containing an antecedent-verifying world-assignment pair for each row separately, as in (59). This yields the high reading.

(59)

| | | | | | |
|---|---|---|---|---|---|
| $g_4$ | | | | | $\langle w_7, g_4\rangle$, $\langle w_8, g_4\rangle$ |
| $g_3$ | | | | $\langle w_5, g_3\rangle$ | |
| $g_2$ | | | $\langle w_3, g_2\rangle$ | $\langle w_5, g_2\rangle$ | |
| $g_1$ | | | $\langle w_2, g_1\rangle$ | | |
| | $w_0$ | $w_1$ | $w_2$, $w_3$, $w_4$ | $w_5$, $w_6$ | $w_7$, $w_8$ |

With these graphical representations in mind, let us now approach the formal implementation suggested by van Rooij (2006). The two selection functions represented by (58) and (59) can be based on defining the similarity orderings (60) and (61) respectively:

(60) $\langle v, h\rangle \leq^{low}_{\langle w,g\rangle} \langle u, k\rangle$ iff $v <_w u$

(61) $\langle v, h\rangle \leq^{high}_{\langle w,g\rangle} \langle u, k\rangle$ iff $h = k \wedge v <_w u$

(60) simply lifts the standard Lewisian similarity relation to world-assignment pairs but changes nothing about the conditions: Pairs are compared based on their respective worlds. $\leq^{low}_{\langle w,g\rangle}$ ranks pairs exactly in the way $<_w$ ranks the worlds of those pairs. In (61), however, we add a second condition: $h = k$, that is, world-assignment pairs can only be ranked with respect to each other if they share an assignment. If they do, they are again ranked based on their worlds, according to $<_w$. This results in a partialization of the standard ordering: We obtain an ordering of pairs for each assignment separately, and the selection function selects the bottom element of all the orderings obtained in this way. This results in quantifying over at least one world-assignment pair for each individual that can be assigned as the referent of an indefinite noun phrase in the antecedent.

In order to generalize these definitions to antecedents with possibly more than one indefinite noun phrase, where each indefinite can obtain either a high or a low reading, we combine (60) and (61) into the lexical entry in (62). In (62), a contextually given set $X$ of variables modulates the behaviour of the similarity relation in the following way: To be compared, world-assignment pairs are required to agree in the values their assignment functions assign to the variables in $X$.

(62) $\langle v,h\rangle \leq^{X}_{\langle w,g\rangle} \langle u,k\rangle$ iff $h \uparrow^X = k \uparrow^X \wedge v <_w u$

For an empty $X$, the requirement $h \uparrow^X = k \uparrow^X$ is vacuous, reducing to the low reading. Adding variables to $X$ partializes the similarity ordering for each such variable, yielding the respective high readings.

Finally, the lexical entries for the counterfactual itself and the selection function, based on the similarity relation defined in (62), are the following[3]:

(63) $[\![\textit{if } p \textit{ would } q]\!]^{\langle w,g\rangle} = 1$ iff $\forall\langle v,h\rangle \in f_{\langle w,g\rangle}(/p/_g) : \langle v,h\rangle \in /q/_g$

[3] These lexical entries are slightly adapted to make them comparable to the definitions of strict and variably strict approaches in (44) and (48). Specifically, (63) is relativized to an input world-assignment pair; instead of returning the output pair (which would be identical to the input), we consider it to be true iff there is such an output. For the DPL-style notations, see the following derivation in 2.2.3.

(64) $/\phi/_g = \{\langle u,k\rangle \mid \exists\langle v,h\rangle \in \{\langle v,h\rangle \mid v \in W \wedge h = g\} : \langle u,k\rangle \in [\![\phi]\!](\langle v,h\rangle)\}$

(65) $f_{\langle w,g\rangle}(/\phi/_g) = \{\langle v,h\rangle \in /\phi/_g : \neg\exists\langle u,k\rangle \in /\phi/_g : \langle u,k\rangle \leq^{X}_{\langle w,g\rangle} \langle v,h\rangle\}$

According to (63), the counterfactual is true relative to a world-assignment pair $\langle w,g\rangle$ if and only if all world-assignment pairs returned by the selection function relative to $\langle w,g\rangle$ and the antecedent also verify the consequent. Verification is defined as in (64): $/\phi/_g$ returns the set of pairs obtained by interpreting $\phi$ relative to the set of all world-assignment pairs that have $g$ as their assignment, while their worlds can be any member of $W$. This ensures that we can assess non-actual worlds. The selection function returns all those world-assignment pairs that verify the supplied antecedent for which there is no other antecedent-verifying pair that is more similar to $\langle w,g\rangle$ according to the $X$-relative similarity relation in (62)[4].

### 2.2.3 A derivation

In order to show the workings of this system, let's redefine all of DPL from (36) to (40) in terms of world-assignment pairs. Note that this is not necessarily a step we can only motivate with respect to counterfactuals. Rather, one of the (intentional) limitations of DPL is that it only dynamically models the anaphoric potential of expressions. However, it evaluates everything with respect to a single world-independent interpretation function. That is, it does not give us any way of modelling the development of the common ground over time, but rather remains static with respect to worlds. For example, if you consider (36), you can see that for $[\![Rt_1...t_n]\!]$, in case there is no tuple of individuals in $F(R)$, the system simply returns the empty set of assignments. If we want a discourse model that does both – i.e. that models how we exclude both worlds and assignments for not corresponding to our shared beliefs –, we need to rewrite DPL in terms of world-assignment pairs anyway. Under the assumption that, in a discourse that goes beyond a single sentence, the output of a formula serves

[4] van Rooij (2006) introduces some further mechanics in order to deal with weak readings. As these are orthogonal to the issues here, we can safely set them aside.

as the input for the next formula, this will then model both the exclusion of assignments and worlds that do not satisfy the requirements of the discourse. The counterfactual then only inhabits a special role in this system with respect to its ability to consider worlds outside the common ground for evaluation.

For this reason, I will first consider a world-assignment based DPL without the counterfactual and show how the derivation of a simple indicative donkey sentence proceeds in this system. I will then add the counterfactual and provide the same derivation for it.

For simplicity, we will assume that the interpretation of individual terms is independent of worlds, i.e. names are constant and variables are resolved only through assignments. However, the interpretation function $F$ will now be world-dependent, i.e. it will return a set of $n$-tuples for an $n$-place predicate and a world, and consequently take an additional world argument.

(66) $[\![Rt_1...t_n]\!] = \{\langle\langle w, g\rangle, \langle v, h\rangle\rangle \mid h = g \wedge w = v \wedge \langle [\![t_1]\!]^h, ...[\![t_n]\!]^h\rangle \in F(R, w)\}$

The other lexical entries are then defined along the same lines as in (37) – (40), simply requiring identity of worlds for the input and output pair in addition to the usual interactions with assignments.

(67) $[\![\exists x\phi]\!] = \{\langle\langle w, g\rangle, \langle v, h\rangle\rangle \mid w = v \wedge \exists k : k[x]g \wedge \langle\langle w, k\rangle, \langle w, h\rangle\rangle \in [\![\phi]\!]\}$

(68) $[\![\phi \wedge \psi]\!] = \{\langle\langle w, g\rangle, \langle v, h\rangle\rangle \mid w = v \wedge \exists k : \langle g, k\rangle \in [\![\phi]\!]$
$\wedge \langle\langle w, k\rangle, \langle w, h\rangle\rangle \in [\![\psi]\!]\}$

(69) $[\![\phi \rightarrow \psi]\!] = \{\langle\langle w, g\rangle, \langle v, h\rangle\rangle \mid w = v \wedge h = g$
$\wedge \forall k : \langle\langle w, h\rangle, \langle w, k\rangle\rangle \in [\![\phi]\!] \rightarrow \exists j : \langle\langle w, k\rangle, \langle w, j\rangle\rangle \in [\![\psi]\!]\}$

(70) $[\![\forall x\phi]\!] = \{\langle\langle w, g\rangle, \langle v, h\rangle\rangle \mid w = v \wedge h = g$
$\wedge \forall k : k[x]g \rightarrow \exists j : \langle\langle w, k\rangle, \langle w, j\rangle\rangle \in [\![\phi]\!]\}$

Since the identity of worlds is always required, we can rewrite this in a slightly simpler form:

(71) $[\![\exists x\phi]\!] = \{\langle\langle w, g\rangle, \langle w, h\rangle\rangle \mid \exists k : k[x]g \wedge \langle\langle w, k\rangle, \langle w, h\rangle\rangle \in [\![\phi]\!]\}$

(72) $[\![\phi \wedge \psi]\!] = \{\langle\langle w, g\rangle, \langle w, h\rangle\rangle \mid \exists k : \langle g, k\rangle \in [\![\phi]\!]$
$\wedge \langle\langle w, k\rangle, \langle w, h\rangle\rangle \in [\![\psi]\!]\}$

(73) $[\![\phi \rightarrow \psi]\!] = \{\langle\langle w, g\rangle, \langle w, h\rangle\rangle \mid h = g$
$\wedge \forall k : \langle\langle w, h\rangle, \langle w, k\rangle\rangle \in [\![\phi]\!] \rightarrow \exists j : \langle\langle w, k\rangle, \langle w, j\rangle\rangle \in [\![\psi]\!]\}$

(74) $[\![\forall x\phi]\!] = \{\langle\langle w, g\rangle, \langle w, h\rangle\rangle \mid h = g$
$\wedge \forall k : k[x]g \rightarrow \exists j : \langle\langle w, k\rangle, \langle w, j\rangle\rangle \in [\![\phi]\!]\}$

The equivalence in (33), repeated below as (75), holds in this system too, as the derivation below shows.

(75) $\exists x Px \rightarrow Qx \equiv \forall x(Px \rightarrow Qx)$

(76) a. $[\![\exists x Px]\!]$
$= \{\langle\langle w, g\rangle, \langle w, h\rangle\rangle \mid \exists k : k[x]g$
$\wedge \langle\langle w, k\rangle, \langle w, h\rangle\rangle \in [\![Px]\!]\}$ by (71)

$= \{\langle\langle w, g\rangle, \langle w, h\rangle\rangle \mid \exists k : k[x]g$
$\wedge w = w \wedge k = h \wedge [\![x]\!]^h \in F(P, w)\}$ by (66)

$= \{\langle\langle w, g\rangle, \langle w, h\rangle\rangle \mid h[x]g \wedge h(x) \in F(P, w)\}$ by $k = h$

b. $[\![\exists x Px \rightarrow Qx]\!]$
$= \{\langle\langle w, g\rangle, \langle w, h\rangle\rangle \mid h = g \wedge \forall k : \langle\langle w, h\rangle, \langle w, k\rangle\rangle$
$\in [\![\exists x Px]\!] \rightarrow \exists j : \langle\langle w, k\rangle, \langle w, j\rangle\rangle \in [\![Qx]\!]$ by (73)

$= \{\langle\langle w, g\rangle, \langle w, h\rangle\rangle \mid h = g \wedge \forall k : (k[x]h \wedge h(x) \in F(P, w))$
$\rightarrow \exists j : \langle\langle w, k\rangle, \langle w, j\rangle\rangle \in [\![Qx]\!]$ by (76a)

$= \{\langle\langle w, g\rangle, \langle w, h\rangle\rangle \mid h = g \wedge \forall k : (k[x]h \wedge h(x) \in F(P, w))$
$\rightarrow \exists j : k = j \wedge j(x) \in F(Q, w)\}$ by (66)

$= \{\langle\langle w, g\rangle, \langle w, h\rangle\rangle \mid h = g \wedge \forall k : (k[x]h \wedge h(x) \in F(P, w))$
$\rightarrow k(x) \in F(Q, w)\}$ by $k = j$

c. $[\![Px \rightarrow Qx]\!]$
$= \{\langle\langle w, g\rangle, \langle w, h\rangle\rangle \mid h = g \wedge \forall k : \langle\langle w, h\rangle, \langle w, k\rangle\rangle$
$\in [\![Px]\!] \rightarrow \exists j : \langle\langle w, k\rangle, \langle w, j\rangle\rangle \in [\![Qx]\!]$ by (73)

$= \{\langle\langle w, g\rangle, \langle w, h\rangle\rangle \mid h = g \wedge \forall k : (h = k \wedge k(x) \in F(P, w))$
$\rightarrow \exists j : (k = j \wedge j(x) \in F(Q, w))\}$ by (66)

$= \{\langle\langle w, g\rangle, \langle w, h\rangle\rangle \mid h = g \wedge ((h(x) \in F(P, w))$
$\rightarrow (h(x) \in F(Q, w)))\}$ by $k = j$ and $h = k$

d. $[\![\forall x(Px \rightarrow Qx)]\!]$
$= \{\langle\langle w, g\rangle, \langle w, h\rangle\rangle \mid h = g \wedge \forall k : k[x]g$
$\rightarrow \exists j : \langle\langle w, k\rangle, \langle w, j\rangle\rangle \in [\![Px \rightarrow Qx]\!]$ by (74)

$= \{\langle\langle w, g\rangle, \langle w, h\rangle\rangle \mid h = g \wedge \forall k : k[x]g$
$\rightarrow \exists j : k = j \wedge ((j(x) \in F(P, w))$
$\rightarrow (j(x) \in F(Q, w)))\}$ by (76c)

$= \{\langle\langle w, g\rangle, \langle w, h\rangle\rangle \mid h = g \wedge \forall k : k[x]g$
$\rightarrow (k(x) \in F(p, w) \rightarrow k(x) \in F(Q, w))\}$ by $k = j$

$= \{\langle\langle w, g\rangle, \langle w, h\rangle\rangle \mid h = g \wedge \forall k : (k[x]g \wedge k(x) \in F(P, w))$
$\rightarrow k(x) \in F(Q, w)\}$ by Exportation

$= \{\langle\langle w, g\rangle, \langle w, h\rangle\rangle \mid h = g \wedge \forall k : (k[x]h \wedge k(x) \in F(P, w)$
$\rightarrow k(x) \in F(Q, w)\}$ by $h = g$

Again, compare the last lines in (76b) and (76d) to see that they are identical, as in the original DPL. Now we can take a look at how the counterfactual interacts with the rest of the system. The DPL-style lexical entry for the counterfactual in (63) can be found in (77) – note that the counterfactual, too, is externally static with respect to both worlds and assignments.

The staticness with respect to worlds models the fact that the counterfactual, while internally accessing worlds outside the common ground, does not add any worlds back into the common ground that have previously been excluded. Rather, we exclude worlds from the common ground if the counterfactual does not hold in them. We inherit the staticness with respect to assignments from DPL, where the (indicative) conditional is equally static with respect to assignments. Whether this is desirable depends on empirical data outside of the scope of this dissertation, and any modifications made to the classical conditional should equally be considered for the counterfactual conditional. However, for the purposes of this dissertation, we will stick to the original DPL.

(77) $[\![\phi > \psi]\!] = \{\langle\langle w, g\rangle, \langle w, g\rangle\rangle \mid \forall\langle v, h\rangle \in f_{\langle w,g\rangle}(/\phi/_g) : \langle v, h\rangle \in /\psi/_g\}$

The $/\cdot/_g$ notation is simply an abbreviation, as spelled out in (64), that allows for an interpretation of a formula not with respect to the original input set of world-assignment pairs, but with respect to a new set, constructed by taking the cross product of $W$, the set of all worlds, and the assignments from the input set. In effect, this allows the counterfactual to range over worlds outside the common ground, but keeps anaphoric relations intact. Consider the following derivation of a counterfactual donkey sentence.

(78) $[\![\exists xPx]\!](\langle v, h\rangle)$
$= \{\langle v, g\rangle \mid g[x]h \wedge g(x) \in F(P, v)\}$ by (76a)

(79) $/\exists xPx/_g$
$= \{\langle u, k\rangle \mid \exists\langle v, h\rangle \in \{\langle v, h\rangle \mid v \in W \wedge h = g\} :$
$\langle u, k\rangle \in [\![\exists xPx]\!](\langle v, h\rangle)\}$ by (64)

$= \{\langle u, k\rangle \mid \exists\langle v, h\rangle \in \{\langle v, h\rangle \mid v \in W \wedge h = g\} :$
$u = v \wedge k[x]h \wedge k(x) \in F(P, v)\}$ by (78)

$= \{\langle u, k\rangle \mid u \in W \wedge k[x]g \wedge k(x) \in F(P, u)\}$ by $u = v, h = g$

(80) $[\![Qx]\!](\langle v,h\rangle)$
$= \{\langle v,h\rangle \mid h(x) \in F(Q,v)\}$ by (66)

(81) $/Qx/_g$
$= \{\langle u,k\rangle \mid \exists\langle v,h\rangle \in \{\langle v,h\rangle \mid v \in W \wedge h = g\} :$
$\langle u,k\rangle \in [\![Qx]\!](\langle v,h\rangle)\}$ by (64)

$= \{\langle u,k\rangle \mid \exists\langle v,h\rangle \in \{\langle v,h\rangle \mid v \in W \wedge h = g\} :$
$u = v \wedge k = h \wedge k(x) \in F(Q,v)\}$ by (80)

$= \{\langle u,k\rangle \mid u \in W \wedge k(x) \in F(Q,u)\}$ by $u = v, k = h, h = g$

(82) $f_{\langle w,g\rangle}(/\exists xPx/_g)$
$= \{\langle v,h\rangle \in /\exists xPx/_g \mid \neg\exists\langle u,k\rangle \in /\exists xPx/_g : \langle u,k\rangle \leq^X_{\langle w,g\rangle} \langle v,h\rangle\}$ by (65)

$= \{\langle v,h\rangle \mid v \in W \wedge h[x]g \wedge h(x) \in F(P,v) \wedge$
$\neg\exists\langle u,k\rangle \in /\exists xPx/_g : \langle u,k\rangle \leq^X_{\langle w,g\rangle} \langle v,h\rangle\}$ by (79)

$= \{\langle v,h\rangle \mid v \in W \wedge h[x]g \wedge h(x) \in F(P,v) \wedge$
$\neg\exists\langle u,k\rangle : u \in W \wedge k[x]g \wedge k(x) \in F(P,u) \wedge$
$\langle u,k\rangle \leq^X_{\langle w,g\rangle} \langle v,h\rangle\}$ by (79)

(83) $[\![\exists xPx > Qx]\!]$
$= \{\langle\langle w,g\rangle,\langle w,g\rangle\rangle \mid \forall\langle v,h\rangle \in f_{\langle w,g\rangle}(/\exists xPx/_g) :$
$\langle v,h\rangle \in /Qx/_g\}$ by (77)

$= \{\langle\langle w,g\rangle,\langle w,g\rangle\rangle \mid \forall\langle v,h\rangle : (v \in W \wedge h[x]g \wedge h(x) \in F(P,v) \wedge$
$\neg\exists\langle u,k\rangle : u \in W \wedge k[x]g \wedge k(x) \in F(P,u) \wedge$
$\langle u,k\rangle \leq^X_{\langle w,g\rangle} \langle v,h\rangle\}) \rightarrow (\langle v,h\rangle \in /Qx/_g)\}$ by (82)

$$= \{\langle\langle w, g\rangle, \langle w, g\rangle\rangle \mid \forall\langle v, h\rangle : (v \in W \wedge h[x]g \wedge h(x) \in F(P, v) \wedge$$
$$\neg\exists\langle u, k\rangle : u \in W \wedge k[x]g \wedge k(x) \in F(P, u) \wedge$$
$$\langle u, k\rangle \leq^{X}_{\langle w,g\rangle} \langle v, h\rangle\}) \rightarrow (v \in W \wedge h(x) \in F(Q, v))\} \qquad \text{by (81)}$$

Admittedly, (83) is not the easiest formula to read, but it does its job. Let's take it apart and see what that job is. For each world-assignment $\langle w, g\rangle$ pair in the input set, the counterfactual conditional does the following: It considers all pairs $\langle v, h\rangle$ which can be constructed from the cross product of $W$ and those assignments that assign an individual with the property $P$ to $x$ in $v$, and for which there is no other pair with the same property that is more similar to the input pair $\langle w, g\rangle$. If all pairs under consideration are such that $x$ also has the property $Q$ in $v$, the original input pair is returned as an output pair. In effect, the counterfactual returns those world-assignment pairs for which the closest world-assignment pairs verify the material conditional, where closeness is either partialized by assignments or not, depending on the variables in $X$.

## 2.3 Dynamic strict counterfactuals

One of the main motivations for van Rooij's (2006) analysis is to account for the licensing of the Negative Polarity Item (NPI) *any* in the antecedent of counterfactual conditionals, as observed by Heim (1984) and demonstrated in (84).

(84) If John owned any donkey, he would beat it.

Specifically, van Rooij argues that his analysis can account for the licensing of *any* in sentences like (84) without making it necessary to move from Lewis's (1973) variably strict analysis to a dynamic strict analysis, as proposed by, e.g., von Fintel (1999, 2001). However, in Walker and Romero (2015), we observe that this way of accounting for the licensing of *any* actually runs into trouble with respect to low readings. In this section, I will

first give a brief introduction to the phenomenon of NPIs, before considering van Rooij's proposal. I will then demonstrate that van Rooij's solution undergenerates in low readings, motivating a move to von Fintel's semantics for the counterfactual. The remainder of the section will describe von Fintel's system and show how van Rooij's insights can be translated into such a dynamic strict framework.

### 2.3.1 Negative Polarity Items

Negative Polarity Items (NPIs) are lexical items with a limited distribution. The archetypical environment in which they are licensed is under negation – hence the name – as demonstrated with the NPI *ever* below:

(85) * I have ever been to Colchester.

(86) I haven't ever been to Colchester.

However, there is a number of other environments in which NPIs are also licensed.

(87) Have you ever been to Colchester?

(88) Only John has ever been to Colchester.

(89) Every student who has ever been to Colchester loved it there.

Various NPIs are known, including the aforemention *ever* and *any*. The "classical" analysis of NPIs is due to Fauconnier (1975) and Ladusaw (1980). It classifies the contexts in which NPIs can appear as downward-monotone environments, with the definition running as follows:

(90) A function $f$ of type $\langle \sigma, \tau \rangle$ is downward-entailing iff for all $x$, $y$ of type $\sigma$ such that $x \Rightarrow y$: $f(y) \Rightarrow f(x)$.

However, this analysis only covers a limited range of the empirically attested cases. Specifically, it does not cover (88) and is unclear on (87). It also leads to the following puzzle:

(91) If John had owned any donkey, he would have beaten it.

The acceptability of (91) is easily explained if we assume that antecedents of counterfactuals are downward-monotone, which is the case for a strict conditional analysis. However, in a variably strict analysis, counterfactual antecedents do not have this property, leaving (91) unexplained. There is obviously a tension between both desiderata, and various theories have been proposed to both cover the remaining cases and resolve this tension. While a detailed consideration of the NPI literature is outside the scope of this dissertation, we will proceed to consider two frameworks: Kadmon and Landman's (1993) widening analysis, which van Rooij (2006) uses, and the dynamic strict semantics proposed by von Fintel (1999, 2001). I will explicitly consider the value of these theories only with respect to the phenomena at hand, leaving other considerations aside.

### 2.3.2 *Any* in high readings

In order to avoid moving from a variably strict to a dynamic strict semantics, van Rooij (2006) suggests to explain the licensing of NPI-*any* in a different way, based on his semantics for counterfactual donkeys. His proposal builds on Kadmon and Landman's (1993) widening analyis of NPIs[5], where *any* is taken to widen the domain of quantification – compared to a "regular" indefinite like *a* or *some* –, and where such a widening – and consequently, the use of *any* – is licensed whenever the resulting proposition is logically stronger than the proposition expressed by an equivalent utterance with a non-widening indefinite expression in place of *any*.

To demonstrate domain widening, Kadmon and Landman present data like the following pair:

(92) a. I don't have potatoes.

b. I don't have any potatoes.

[5] Note that independently of the discussion in this chapter, there are arguments against this analysis, e.g. Arregui (2008).

Their claim is that (92a) is acceptable even if there are exceptions – say, if we are cooking and I have only rotten potatoes in my basement. In (92b), on the other hand, the use of *any* signals a wider domain of quantification, including even the potatoes that are not strictly relevant for cooking.

They then suggest that the use of *any* is only licensed if its use strengthens the utterance, that is, if the resulting utterance is logically stronger than the utterance with a simple indefinite. It is easy to see how this explains the negation data from (85) and (86): In a positive environment, making an existential statement over a larger domain actually weakens the utterance, whereas in a negative environment, the resulting utterance is logically stronger – not only do we deny the existence of relevant potatoes, but the existence of potatoes at all.

Clearly, this reasoning also holds for the high reading of counterfactual donkey sentences under van Rooij's (2006) analysis. Consider the following pair of sentences, with the domain of quantification for the indefinite expressions *a* and *any* indicated as subscripts:

(93) If John had owned $\text{a}_{\{d_1,d_2\}}$ donkey, he would have beaten it.

(94) If John had owned $\text{any}_{\{d_1,d_2,d_3\}}$ donkey, he would have beaten it.

Given the analysis in 2.2.2, (94) entails (93): Where (93) amounts to the conjunction of (95)–(96), (94) is the conjunction of (95)–(97) and therefore logically stronger, as required in Kadmon and Landman's (1993) analysis.

(95) If John had owned $\text{d}_1$, John would have beaten $\text{d}_1$.

(96) If John had owned $\text{d}_2$, John would have beaten $\text{d}_2$.

(97) If John had owned $\text{d}_3$, John would have beaten $\text{d}_3$.

It is correct that for these cases, van Rooij's (2006) machinery for high readings explains the licensing of NPI-*any*, without necessitating a move to a dynamic strict semantics.

### 2.3.3 *Any* in low readings

However, this analysis does not successfully carry over to counterfactuals under a low reading. First, it can be established that NPI-*any* is also licensed in the antecedents of low reading counterfactuals by employing an identificational sentence, which cannot have a high reading:

(98) If John had owned any donkey, it would have been Platero.

However, as a comparison between the following sentence pair shows, widening the domain of quantification does not lead to a logically stronger sentence.

(99) If John had owned $a_{\{d_1,d_2\}}$ donkey, it would have been Platero$^{d_1}$.

(100) If John had owned $any_{\{d_1,d_2,d_3\}}$ donkey, it would have been Platero$^{d_1}$.

In fact, the semantics of the low reading is exactly the semantics of the standard Lewis-Stalnaker analysis, which – through rejecting a fixed domain of quantification – renders the counterfactual non-downward-entailing. So it is not surprising that van Rooij's (2006) semantics, which leaves low readings untouched, cannot account for NPI-licensing there. Thus, the acceptability of (98) shows that we need an account of NPI-licensing that is independent of van Rooij's additional mechanisms which only apply to high readings.

### 2.3.4 The modal horizon

The standard Ladusaw-Fauconnier analysis of Negative Polarity Items (Fauconnier, 1975, 1978; Ladusaw, 1980) proposes that NPIs are licensed in downward-entailing contexts. This explains much of their distribution; however, it famously leaves their acceptability in the antecedents of counterfactual conditionals unexplained. In contrast to the strict conditional analysis, the Lewis-Stalnaker analysis (Lewis, 1973b; Stalnaker, 1968) does not verify strengthening of the antecedent and therefore does not ensure that antecedents are downward-entailing. For this reason, von Fintel (1999,

2001) proposes an adaptation of both analyses in two steps: First, adjusting the definition of downward-entailingness used in the Ladusaw-Fauconnier analysis, and second, moving from the variably strict Lewis-Stalnaker semantics to a version of the strict conditional semantics that preserves the advantages of Lewis-Stalnaker sketched in Section 2.2 while at the same time being downward-entailing in the adjusted sense.

Standardly, downward-entailingness (DE) is defined as follows (von Fintel, 1999):

(101) A function $f$ of type $\langle \sigma, \tau \rangle$ is downward-entailing (DE) iff for all $x$, $y$ of type $\sigma$ such that $x \Rightarrow y : f(y) \Rightarrow f(x)$

Von Fintel calls his adjusted definition *Strawson*-downward-entailingness (SDE), and defines it as follows:

(102) A function $f$ of type $\langle \sigma, \tau \rangle$ is Strawson-downward-entailing (SDE) iff for all $x$, $y$ of type $\sigma$ such that $x \Rightarrow y$ and $f(x)$ is defined: $f(y) \Rightarrow f(x)$

The difference between (101) and (102) is in the additional condition that $f(x)$ be defined. This weakens the notion of downward-entailingness, as cases in which this additional condition does not hold do not count against a function being SDE, that is, a larger class of environments will be SDE than DE. One application for this weakened definition is in explaining the licensing of NPIs under *only*, as in (88). Consider the following sentence:

(103) Only John reads a book.

Under the classical definition of downward-entailingness in (101), this is not downward-entailing, as it does not allow to infer (104):

(104) Only John reads *Grisella*.

Why can we not infer (104) from (103)? Because it might well be true that no one but John reads a book – verifying (103) –, but not that John reads *Grisella*. However, under von Fintel's adapted notion of Strawson-DE, we are allowed to draw the inference by adding the additional premise that the presupposition $\pi$ of (104) be defined[6]:

(105) $\pi$ = John reads Grisella.

(106) Only John reads a book $\wedge \pi \Rightarrow$ Only John reads Grisella

The second step is to ensure that we have an analysis of the counterfactual that makes antecedents SDE. The crucial difference between a strict semantics (which is DE *simpliciter*) and a variably strict semantics is in the fixedness of the domain of quantification. Von Fintel proposes a middle position between the two: In a dynamic strict semantics, the domain is quasi-fixed, but can dynamically be updated through a process of presupposition accommodation. Since SDE only requires monotonicity under the assumption of fulfilled presuppositions, this accommodation is invisible to it, rendering the antecedents of counterfactuals SDE.

In terms of semantics for the counterfactual, we return to the strict picture:

(107) $[\![\textit{if } p \textit{ would } q]\!]^w = 1$ iff $\forall w' \in D_w : w' \in [\![p]\!] \rightarrow w' \in [\![q]\!]$

However, there is a twist to the strict account: We now take $D_w$ to be a dynamic object, the *modal horizon* of $w$, which evolves throughout discourse. It is primarily updated through a presupposition that is attached to the definition in (107):

(108) $[\![\textit{if } p \textit{ would } q]\!]^w$ is defined only if $[\![\phi]\!] \cap D_w \neq \emptyset$

[6] As with the Kadmon and Landman (1993) analysis, there are independent arguments against this analysis that we will set aside for the remainder of this discussion. Note that the analysis of the prejacent as a presupposition of *only* employed in this argument is not uncontroversial, see e.g. Ippolito (2006, 2008) for an overview.

That is, a counterfactual conditional is only defined if there are antecedent-worlds in the modal horizon. If the modal horizon does not contain antecedent worlds, i.e. the presupposition fails, it is accommodated by expanding the horizon minimally to include at least one antecedent world. This expansion follows the rule in (109)[7]:

(109) $\phi$-expanded $D_w = D_w \cup \{w' \mid \forall v \in [\![\phi]\!] : w' \leq_w v\}$

That is, the modal horizon is expanded to include any world which is closer or equally close to the actual world $w$ than all antecedent-worlds. This includes all worlds up to (and including) the closest antecedent-worlds. In effect, this is entirely equivalent to the variably strict reading in those cases where the modal horizon did not previously include antecedent-worlds[8].

However, in contrast to the variably strict semantics, this semantics is SDE, providing an explanation for the licensing of NPI-*any* in the antecedent: For any sentences $\phi$, $\psi$ and $\tau$, and any domain $D_w$ for which both $[\![\phi \boxright \psi]\!]$ and $[\![(\phi \wedge \tau) \boxright \psi]\!]$ are defined, the former will entail the latter. This is guaranteed because a domain $D_w$ that is defined for the antecedent $(\phi \wedge \tau)$ will ensure that there are $\tau$-worlds amongst the $\phi$-worlds; consequently, if $\psi$ holds throughout all $\phi$-worlds, it will also hold throughout those $(\phi \wedge \tau)$-worlds. The variably strict semantics does not achieve this, because it cannot guarantee that the $(\phi \wedge \tau)$-worlds of the second counterfactual are included in the $\phi$-worlds the first counterfactual is quantifying over.

[7] Note that the modal horizon is defined relative to a world, as in von Fintel (1999, 2001), a detail that is somewhat obscured in Walker and Romero (2015), where the subscript $w$ is not shown, while our update procedure in that paper still clearly assumes a dependency on world-assignment pairs.

[8] In cases where the modal horizon already includes antecedent-worlds, there is no guarantee that these are the closest ones. This leads to the desired behaviour in so-called reverse Sobel-sequences, an additional motivation for von Fintel's (1999) account that is, however, orthogonal to the issues here.

### 2.3.5 A dynamic strict conditional analysis

Given the data in 2.3.3 and von Fintel's (1999) proposal, Walker and Romero (2015) argue that moving to a dynamic strict analysis remains necessary. However, as von Fintel's analysis is not aimed at explaining the behaviour of indefinites, it does not address the issue of high readings in counterfactual donkey sentences. The best option seems to be a combination of both approaches, combining DPL with a dynamic strict semantics and adding some of van Rooij's machinery.

Let us begin by setting up a dynamic strict analysis in a DPL-style framework first, before adding the necessary machinery for obtaining high readings.

(110) $[\![\textit{if } p \textit{ would } q]\!]^{\langle w,g\rangle} = 1$ iff $\forall\langle v,h\rangle \in D_{\langle w,g\rangle} : \langle v,h\rangle \in /p/_g \rightarrow \langle v,h\rangle \in /q/_g$

In parallel to the difference between the variably strict and the dynamic strict semantics in (48) and (107), our semantics in (110) are equivalent to the variably strict semantics in (63), with the only difference being that the set of world-assignment pairs to be quantified over is now given by a (quasi-fixed) domain $D_{\langle w,g\rangle}$, rather than the selection function $f_{\langle w,g\rangle}$. In order to guarantee SDE, we make the same move as von Fintel (1999), requiring $D_{\langle w,g\rangle}$ to contain at least one antecedent-world for the counterfactual to be defined, and providing an update procedure for the modal horizon if that presupposition is violated. However, rather than having a modal horizon relative to a world $w$, we now define the modal horizon as relative to a world-assignment pair $\langle w,g\rangle$, which amounts to a set of modal horizons, one per assignment. The challenge is now to provide an adequate update procedure for this set[9].

The difference between a variably strict approach and a modal horizon approach is the latter's permanency. While the variably strict selection function – as defined in (65) – only needs to return those world-assignment pairs we immediately want to quantify over, in the case of a modal horizon,

[9] Note that the account set out here, although building on our ideas from Walker and Romero (2015), is a more developed version that allows for a clearer exposition and resolves some presentational issues. Specifically, the update in Walker and Romero (2015) produces a "gappy" modal horizon, as opposed to the analysis outlined below.

we need to develop an object with more internal consistency. In particular, modal horizons are not gappy, that is, they always extend outwards to the antecedent-worlds they need to include by also including all less similar worlds. In defining an update procedure for a modal horizon, we need to take this property into account.

However, other than that, the strength of the dynamic strict approach is exactly its similarity to the variably strict approach. This allows us to make use of our previous definitions (64) and (65), repeated below as (111) and (112):

(111) $/\phi/_g = [\![\phi]\!](\{\langle v, h\rangle \mid v \in W \wedge h = g\})$

(112) $f_{\langle w,g\rangle}(/\phi/_g) = \{\langle v, h\rangle \in /\phi/_g : \neg\exists\langle u, k\rangle \in /\phi/_g : \langle u, k\rangle \leq^{X}_{\langle w,g\rangle} \langle v, h\rangle\}$

The worlds returned by $f_{\langle w,g\rangle}$ form what I will call the *border elements* of our modal horizon update[10]. We will then need to expand the modal horizon up to these elements. The formula in (113), which we will have to revise accordingly, gives us a "gappy" modal horizon that only includes the border elements themselves:

(113) $\phi$-expanded $D_{\langle w,g\rangle} = D_{\langle w,g\rangle} \cup \{\langle w', g'\rangle \mid \langle w', g'\rangle \in f_{\langle w,g\rangle}(/\phi/_g)\}$

In order to fill the gaps, we will have to add all world-assignment pairs that share an assignment with a border element, but have as their world a world that is closer to $w$ than the border element's world (or equally close, in order to also include the border elements themselves):

(114) $\phi$-expanded $D_{\langle w,g\rangle} = D_{\langle w,g\rangle} \cup \{\langle w', g'\rangle \mid \exists\langle v, h\rangle \in f_{\langle w,g\rangle}(/\phi/_g) : h = g' \wedge w' \leq_w v\}$

According to the revised (114), a pair $\langle w', g'\rangle$ is added to the modal horizon if there is a border element $\langle v, h\rangle$ such that $h = g'$ (i.e. both share an assignment) and for which $w'$ is closer to $w$ than the border element's $v$.

[10] Note that they will also have to play a role in the presupposition attached to the counterfactual: If we want to ensure that an available antecedent-world does not block a high reading, we need to require that the modal horizon includes the full set of border elements. The selection function $f_{\langle w,g\rangle}$ will then also modulate between different presuppositions for high and low counterfactuals.

### 2.3.6 Application: Generating high and low readings

In order to demonstrate the working system, we will apply the definitions from (110)–(112) and (114) to an example. For the sentence in (115), and the model described in (116), I will first show how to derive a high, and then how to derive a low reading.

(115) If John had owned a$^x$ donkey, he would have beaten it$_x$.

(116) John does not own or beat any donkeys. He is most likely to own donkey *a*, and he is very likely to beat it. He is less likely to own either donkey *b* or *c* and would most likely not beat them.

(117)

| | owned-by-John | beaten-by-John |
|---|---|---|
| $w_0$ | $\emptyset$ | $\emptyset$ |
| $w_1$ | {a} | {a} |
| $w_2$ | {a} | $\emptyset$ |
| $w_3$ | {b} | $\emptyset$ |
| $w_4$ | {c} | $\emptyset$ |

(118) $w_0 <_{w_0} w_1 <_{w_0} w_2 <_{w_0} w_3, w_4$

The sentence in (115) can be judged as either true – under a low reading – or false – under a high reading. We will begin by generating the (true) low reading. Assuming a discourse-initial empty assignment function $g$ and empty modal horizon $D_{\langle w_0,g\rangle}$, we interpret (115) as follows, according to (110):

(119) $[\![$*if [John owned a$^x$ donkey], would [he beat it$_x$]*$]\!]^{\langle w_0,g\rangle}$
$= 1$ iff $\forall\langle v,h\rangle \in D_{\langle w_0,g\rangle}$ :
$\langle v,h\rangle \in$ /John owned a$^x$ donkey/$_g$ $\rightarrow \langle v,h\rangle \in$ /he beat it$_x$/$_g$

First, we need to check whether $D_{\langle w_0,g\rangle}$ contains any antecedent-worlds. This is obviously not the case, as it is empty. So we need our accommodation procedure, as defined in (114):

(120) Expanded $D_{\langle w_0,g\rangle}$
$= \emptyset \cup \{\langle w', g'\rangle \mid \exists\langle v,h\rangle \in f_{\langle w_0,g\rangle}$(/John owned a$^x$ donkey/$_g$) :
$h = g' \wedge w' \leq_{w_0} v\}$

The update procedure in turn requires us to identify the border elements according to $f_{\langle w_0,g\rangle}$ applied to the antecedent. Using the definition in (112), we obtain the following:

(121) $f_{\langle w_0,g\rangle}(/\text{John owned a}^x\ \text{donkey}/_g) =$
$\{\langle v,h\rangle \in /\text{John owned a}^x\ \text{donkey}/_g :$
$\neg\exists\langle u,k\rangle \in /\text{John owned a}^x\ \text{donkey}/_g : \langle u,k\rangle \overset{X}{\leq_{\langle w_0,g\rangle}} \langle v,h\rangle\}$

Since we are calculating the low reading, $X = \emptyset$. That is, the comparison between world-assignment pairs $\langle u,k\rangle$ and $\langle v,h\rangle$ happens solely on the basis of their worlds, without considering assignments. Out of the antecedent-world-assignment pairs, we therefore only obtain those that have the world closest to $w_0$ as their world. In the model in (116), this means we obtain only $\langle w_1, g^{a/x}\rangle$. Feeding this border element back into (120) gives us two pairs for the modal horizon: the border element itself, and the pair $\langle w_0, g^{a/x}\rangle$, to avoid gaps. (119) then quantifies over these two pairs.

Despite adding both pairs to the modal horizon, only one of the pairs in the modal horizon – $\langle w_1, g^{a/x}\rangle$ – satisfies the antecedent and therefore matters to the truth conditions of the counterfactual. We require this pair to also satisfy the consequent. Since it is true that John beats $a$ at $w_1$, the counterfactual comes out as true, as required.

The high reading plays out in a similar way, but with $X = \{x\}$. This means that in (121), we compare world-assignment pairs only if their assignments agree in $x$. For each assignment, we then obtain the closest antecedent-world, yielding $\langle w_1, g^{a/x}\rangle$, $\langle w_3, g^{b/x}\rangle$ and $\langle w_4, g^{c/x}\rangle$. Since these are only the border elements, we need to fill the gaps, adding $\langle w_0, g^{a/x}\rangle$, $\langle w_0, g^{b/x}\rangle$, $\langle w_1, g^{b/x}\rangle$, $\langle w_2, g^{b/x}\rangle$, $\langle w_0, g^{c/x}\rangle$, $\langle w_1, g^{c/x}\rangle$, $\langle w_2, g^{c/x}\rangle$ and $\langle w_3, g^{c/x}\rangle$. Quantification, again, is then restricted to the antecedent-pairs, i.e. the border elements in our case. Since two of those pairs do not satisfy the consequent, the counterfactual comes out as false, as intended.

## 2.4 Summary and discussion

In this chapter, I have shown the dynamic approach to counterfactual donkey sentences. Combining Dynamic Predicate Logic (Groenendijk and Stokhof, 1991) with Lewis's (1973) variably strict semantics, as proposed by van Rooij (2006), generates both readings, but inherits the variably strict semantics' difficulties in explaining the licensing of NPI-*any* in the antecedents of counterfactual conditionals. The solution proposed by van Rooij (2006) undergenerates, only successfully applying to high but not to low readings. Following Walker and Romero (2015), I therefore develop an account of counterfactual donkey sentences that combines DPL with von Fintel's (2001) dynamic strict semantics.

This account generates high and low readings, given the correct parameter settings for each, that is, depending on the contextually determined variable $X$ that the similarity ordering in the selection function is sensitive to. However, it does not explain the distribution of high and low readings observed in Ch. 1. In order to arrive at a full account of high and low readings, the dynamic account would at least need to be supplemented by a pragmatic account that predicts under which circumstances variables are or are not included in $X$.

One crucial feature of the account presented here is the way in which the (dynamic) semantics of the indefinite (and other quantifiers) and the semantics of the counterfactual are tightly integrated with one another. $X$ is a subset of the variables introduced by quantifiers in the antecedent, and the partialization of similarity in the two-dimensional picture splits the ordering of worlds along assignment lines. This can be seen as a strength of the account, given the strong independent evidence we have for a dynamic analyis of the indefinite in the tradition of Groenendijk and Stokhof (1991). However, this strength is turned into a weakness if it turns out that we want a different analysis of the puzzles solved in the dynamic tradition. The account as presented here would force us to commit us to an analysis that is not without alternatives that should at least be considered.

One clear candidate for such an alternative account is D-type theory in its modern form (Elbourne, 2005, 2013; Heim, 1990). In the next chapter, we are going to consider how the insights from van Rooij's (2006) and this chapter could be translated into the framework of D-type theory.

One potential further argument for not tying our analysis of high and low readings to the semantics of the indefinite comes from cases where no overt indefinite is present in the counterfactual, and yet we get a clear distinction between both types of reading[11]:

(122) If we had gone to Paris, I would have been happy.

The most salient reading of (122) is one where any sufficiently salient way of going to Paris would have made me happy, e.g. if we had gone by airplane, by train or by car (i.e. a high reading over modes of transportation.) However, consider the following:

(123) *Taking the train from Berlin to Moscow.* I always wanted to ride the train from Berlin to Paris. If we had gone to Paris, I would have been happy.

In (123), it is very clear that we are only quantifying over the most likely worlds in which we go to Paris by train. Committing to an analysis as sketched in this chapter also commits us to postulate some assignment-updating operator, e.g. a covert indefinite, in the antecedent of both (122) and (123). However, it seems like the actual high and low readings we obtain are highly contextually dependent, calling into question an analysis that relies on the presence of syntactic operators:

(124) *My friend Pina went to a summer school in Paris, but you and I went to Moscow.* If we had gone to Paris, I would have been happy.

[11] A related point is independently made by Arregui (2005b) and Nichols (2016), but see also AnderBois (2011) whose argument for the presence of existential quantification over events to explain cases of sluicing without an overt indefinite might carry over to our phenomenon.

Here, (124) seems to be dependent on us most likely going to Paris as participants in the same summer school as Pina, not on any particular mode of transportation. However, it is not obvious that we want to introduce indefinites over modes of transportations, modes of attendance, and whatever other scenarios we can come up with. This seems to point towards an account that is formally independent of the indefinite. I return to this point in Ch. 4, where I discuss an alternative account that does not rely on the indefinite.

# Chapter 3

# D-type theory

In the attempt to avoid a theoretical commitment to dynamic semantics, one alternative – and historically precedent – line has often been construed as a major contender in explaining donkey sentences, first under the label of E-type, and later D-type theory. Where dynamic semantics attempt to solve the donkey puzzle by modifying the semantics of predicate logic, and consequently the semantics of the indefinite article – which in dynamic semantics acquires the ability to make lasting changes to the context –, D-type theory shifts the locus of analysis to the other potential culprit in donkey sentences: the pronoun in the consequent.

In this chapter, I will first sketch a modern D-type theory based on the account by Elbourne (2005, 2013). This theory, building on Heim's (1990) re-implementation of E-type theory in a situation semantics framework, argues for a syntactically motivated account in which pronouns spell out definite descriptions in LF, while indefinites introduce minimal situations containing unique referents for these definite descriptions. Minimal situations, construed as consisting of "thin" individuals (Kratzer, 1989) and those of their properties that are introduced by the antecedent, take on a similar role to assignments in dynamic semantics. Specifically, they allow the conditional to quantify over donkey-farmer pairs in the same way as the standard "unselective binding" approach that is implemented in DPL.

However, this implementation is not designed to handle various problems discussed in later dynamic semantics literature, specifically the problems of weak and proportional readings. I will propose some potential amendments that attempt to close this gap based on my previous work in Walker (2014). However, as it turns out, there are limitations inherent to situation semantics that make D-type theory less powerful than dynamic semantics. I will address these shortcomings and how they affect the solution proposed in Walker (2014), before further discussing Elbourne's (2013) extension to counterfactuals. As Elbourne's proposal does not address high readings, I will then sketch a potential solution that recreates van Rooij's (2006) mechanics in D-type theory, before discussing how an approach along these lines runs into similar problems due to the same limitations. The chapter concludes with an assessment of such strategies in general and prepares the argument for the following chapters: In explaining the puzzle of counterfactual donkey sentences, a situation-based account that only seeks to recreate the dynamic solutions sketched in Ch. 2 could at best reach as far as they do, but will in most cases yield even less due to the limitations inherent in D-type theory. In order to construct a theory that explains the data at hand without being limited to the results in the previous chapter, we will have to examine more closely the role that situations could play independently of their use in simulating assignments.

## 3.1 Standard D-type theory

The basic idea of E-type theory was first formulated in Cooper (1979) and Evans (1977, 1980): Pronouns are sometimes (or always) not to be interpreted as (bound or free) variables, but rather have the status of something akin to definite descriptions[1]. Elbourne (2005) further distinguishes between E- and D-type based on the precise implementation of this idea: E-type theories – most prominently represented by Evans (1977, 1980) – assume that pronouns are rigidly referring expressions, that is, they pick out a particular individual across worlds, based on some descriptive content provided by context. D-type theories, on the other hand, – which Elbourne takes to be represented by theories in the tradition of Cooper (1979) and Heim (1990) – actually treat pronouns as definite descriptions, either just semantically or even at the syntactic level. Elbourne himself argues for a syntactic D-type theory in which the pronoun is simply a phonological variant of the definite article, with the descriptive content syntactically present but phonologically elided. That is, the sentence in (125) – our well-known donkey sentence – is taken to be syntactically equivalent to the sentence in (126) at LF:

(125) If a farmer owns a donkey, he beats it.

(126) If a farmer owns a donkey, the farmer beats the donkey.

In this chapter, we will be following Elbourne's implementation. As the most developed D-type theory at this point, it is a strong contender for dynamic semantics, and has explicitly been extended to cover counterfactual conditionals as well as indicative ones (Elbourne, 2013), making it particularly relevant to our puzzle. Many of the arguments made here could

[1] This is also supported by syntactic analyses (Déchaine and Wiltschko, 2002; Patel-Grosz and Grosz, 2017), although Déchaine and Wiltschko suggest a more fine-grained typology of pronouns, where only some of them are full Determiner Phrases. See also Schwarz (2009), who argues that for German, it may be necessary to analyze some pronouns with dynamic semantics, and others with D-type theory. For simplicity, I will assume a unified semantics for pronouns here – as Elbourne (2005) does implicitly –, but these results clearly warrant further research.

probably be carried over to other variants of E- or D-type theory (Büring, 2004; Heim, 1990; Schwarz, 2009); however, for simplicity and concreteness we will limit ourselves to his account as developed in Elbourne (2005, 2013).

Elbourne argues for this particular instantiation of D-type theory based on several data points. The following examples, due to Postal (1966), demonstrate that pronouns can take the syntactic position of a determiner, with an overt descriptive noun phrase:

(127) You troops will embark but the other troops will remain.

(128) We Americans distrust you Europeans.

This can also be demonstrated for German, although it is unclear whether this particular phenomenon generalizes crosslinguistically:

(129) Ihr Truppen werdet ausrücken, aber die anderen Truppen bleiben zurück.

(130) Wir Amerikaner misstrauen euch Europäern.

Note that these examples are intonationally distinct from appositive constructions like the following:

(131) You, troops, will embark.

Assuming that pronouns do indeed take the position of determiners, analyzing pronouns as definite descriptions with their descriptive content standardly elided becomes plausible given data on NP deletion like the following:

(132) Bill's story about Sue may be amazing, but Max's is virtually incredible.

(133) I like Bill's wine, but Max's is even better.

The examples in (132) and (133) show that it is indeed possible to elide noun phrases in the presence of a linguistic antecedent. Taking the possibility of NP-deletion and pronouns in determiner position together, Elbourne then argues for a semantics in which the pronoun does in fact have exactly the semantics of the definite article.

The drawback of a D-type theory, however, is based exactly in the attractive parallel between the pronoun and the definite article: The definite article comes with its own set of problems and debates, one of them about the presuppositions it carries. Definite articles are standardly taken to have a uniqueness presupposition, which generates unwelcome results in donkey sentences (Heim, 1982):

(134) If a man is in Athens, he is not in Rhodes.

If *he* is interpreted as *the man in Athens* and the conditional universally quantifies over worlds or times, Heim argues, we yield the prediction that there is only one unique man in Athens at any given world or time. This is clearly not expressed by the sentence in (134). However, there is a way of circumventing this problem, suggested by Heim (1990) and adopted by Elbourne (2005). Given the tools from Kratzer's (1989) situation semantics, we can let the conditional universally quantify over situations that are smaller than worlds; precisely: over minimal situations that are just large enough to satisfy the uniqueness presuppositions of the definite article.

Kratzer's situation semantics – first proposed in Kratzer (1989), but see also later versions in Kratzer (2012, 2016) – assumes an ontology in which possible worlds are not primitive entities, but rather special (i.e. maximal) cases of another primitive entity, the situation. Situations are composed of individuals and their properties – indeed, an individual and all of its properties constitute just another special case of a situation –, and are ordered by a part-of relation $\sqsubseteq$[2]. Since every situation can be part of at most one world – like individuals, as those are also a special case of situations (Lewis, 1968) –, worlds are maximal situations that are themselves not included in other situations. But situations can be much smaller than worlds – they can, for example, consist of two individuals,

[2] Elbourne himself uses $\leq$ to denote the part-of relation, but since we already employ this symbol to denote the Lewisian similarity relation, to avoid confusion we will instead use $\sqsubseteq$.

a farmer and a donkey, and the owning-relationship that holds between them, while other parts of the world – e.g., most crucially other farmers and donkeys, but also other relations and properties in which the two individuals participate – are excluded.

Assuming this ontology, Elbourne (2005) proposes the following semantics. Note that Elbourne explicitly restricts himself to situations that are part of the actual world, i.e. the semantics below is not designed to deal with possible worlds. We will return to this point later, when we discuss Elbourne's (2013) extension to counterfactuals.

(135) $[\![Mary]\!]^g = \lambda s.Mary$

(136) $[\![owns]\!]^g = \lambda u_{\langle s,e\rangle}.\lambda v_{\langle s,e\rangle}.\lambda s.v(s)$ owns $u(s)$ in $s$

(137) $[\![donkey]\!]^g = \lambda u_{\langle s,e\rangle}.\lambda s.u(s)$ is a donkey in $s$

(138) $[\![a]\!]^g = \lambda f_{\langle\langle s,e\rangle,\langle s,t\rangle\rangle}.\lambda g_{\langle\langle s,e\rangle,\langle s,t\rangle\rangle}.\lambda s.$ there is an individual $x$ and a situation $s'$ such that $s'$ is a minimal situation such that $s' \sqsubseteq s$ and $f(\lambda s.x)(s') = 1$, such that there is a situation $s''$ such that $s'' \sqsubseteq s$ and $s''$ is a minimal situation such that $s' \leq s''$ and $g(\lambda s.x)(s'') = 1$

(139) $[\![it]\!]^g = [\![the]\!]^g = \lambda f_{\langle\langle s,e\rangle,\langle s,t\rangle\rangle}.\lambda s.\iota x f(\lambda s'.x)(s) = 1$ [3]

(140) $[\![always]\!]^g = \lambda p_{\langle s,t\rangle}.\lambda q_{\langle s,t\rangle}.\lambda s.$ for every minimal situation $s'$ such that $s' \sqsubseteq s$ and $p(s') = 1$, there is a situation $s''$ such that $s'' \sqsubseteq s$ and $s''$ is a minimal situation such that $s' \sqsubseteq s''$ and $q(s'') = 1$

(141) $[\![if]\!]^g = \lambda p_{\langle s,t\rangle}.p$

Additionally, Elbourne assumes the LF in (143) for (142):

(142) If a farmer owns a donkey, he beats it.

(143) [[always [if [[a man] [ $\lambda_6$ [[a donkey] [$\lambda_2$ [$t_6$ owns $t_2$]]]]]]] [[he farmer] beats [it donkey]]]

[3] Elbourne explicitly spells out the presupposition $\exists!x f(\lambda s'.x)(s) = 1$ of the $\iota$-operator in his lexical entry, which is redundant. I am dropping it here.

In understanding these lexical entries and the resulting derivation, it is helpful to draw out the parallels to dynamic semantics. Note that where the indefinite in dynamic semantics served to modify assignments – in effect creating one modified assignment for every individual that satisfies the noun phrase associated with the indefinite –, the indefinite in Elbourne's system, i.e. (138), sets up one minimal situation that contains an individual and its property of satisfying the associated noun phrase per such individual. That is, assignments and minimal situations play very similar roles in the two systems. Similarly, Elbourne shares the assumption that the effect of universal quantification does not come from the indefinite itself, but is provided by something external to it. While dynamic semantics places it in the semantics of the conditional itself, he packages this meaning in the lexical entry for the quantificational adverb associated with the conditional, in our case *always* in (140), following Lewis (1975) and Kratzer (1979).

Computing (143) with the semantics in (135) – (141) yields a complex structure of nested situations. The antecedent sets up a situation containing a farmer that owns a donkey; at the same time, this situation is further structured into subsituations: crucially, one containing the (unique) man and one containing the (unique) donkey. The conditional then asserts that for each minimal situation that contains a farmer owning a donkey, there is a corresponding minimal situation that contains the first situation and extends it with the farmer beating the donkey. The minimality requirements ensure the identity of farmer and donkey respectively, as the second situation contains the first, i.e. it also contains its farmer and donkey, and adding any more farmers and donkeys would violate minimality.

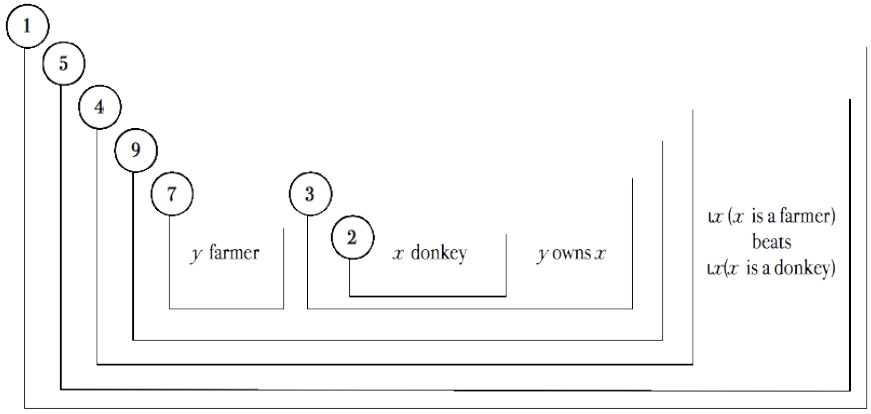
Spelled out, the truth conditions sketched above read as follows:

(144) $\lambda s_1$. for every minimal situation $s_4$ such that
$s_4 \sqsubseteq s_1$ and there is an individual $y$ and a situation $s_7$ such that $s_7$ is a minimal situation such that $s_7 \sqsubseteq s_4$ and $y$ is a farmer in $s_7$, such that there is a situation $s_9$ such that $s_9 \sqsubseteq s_4$ and $s_9$ is a minimal situation such that

$s_7 \sqsubseteq s_9$ and there is an individual $x$ and a situation $s_2$ such that $s_2$ is a minimal situation such that $s_2 \sqsubseteq s_9$ and $x$ is a donkey in $s_2$, such that there is a situation $s_3$ such that $s_3 \sqsubseteq s_9$ and $s_3$ is a minimal situation such that $s_2 \sqsubseteq s_3$ and $y$ owns $x$ in $s_3$,
there is a situation $s_5$ such that

$s_5 \sqsubseteq s_1$ and $s_5$ is a minimal situation such that $s_4 \sqsubseteq s_5$ and $\iota x(x$ is a man in $s_5)$ beats in $s_5$ $\iota x(x$ is a donkey in $s_5)$

(145) *Graphical representation of (144):*




As the graphical representation in (145) shows, the conditional really expresses a relationship between two sets of situations: the situations that are labeled $s_4$ which express the content of the antecedent, and the situations that are labeled $s_5$ and express both the content of the antecedent and the consequent. The conditional is true if and only if for every situation $s_4$, there is a corresponding situation $s_5$.

These truth conditions, although spelled out in terms of situations rather than assignments, turn out to be roughly equivalent to those produced by a dynamic account as described in Ch. 2. There are some border cases where they do not produce the same results unless some further assumptions are made, the so-called "bishop cases", also known as the

problem of indistinguishable participants. The problem in these cases, illustrated by (146), is that even within the minimal antecedent-verifying situations there is no unique referent for the definite description in the consequent.

(146) If a bishop meets a bishop, he greets him.

The problem is not trivial, but as a puzzle to be solved by D-type theory independently of counterfactuals remains orthogonal to the remaining issues discussed in this chapter. We will set it aside for this reason and refer the interested reader to the literature (Barker and Shan, 2008; Elbourne, 2005, 2009, 2010; Kroll, 2008).

## 3.2 The proportion problem

There are a number of classical problems for theories of donkey sentences that can be used to assess their range. In addition to covering the universal reading in the Geach-case in (125), they should also be able to explain the following readings:

(147) If a man has a dime, he throws it into the meter.

(148) If a man has a credit card, he uses it to pay for his meal.

(149) If a farmer owns a donkey, he is usually rich.

(147) and (148) deviate from the standard donkey case in that they do not require all men to throw all their dimes into the parking meter, or to use all their credit cards to pay for each meal; rather, they require all men to throw one of their dimes into the meter, and to pay their meals with one of their credit cards respectively. That is, the quantificational force of the second indefinite here is existential rather than universal. This reading is usually known as the *weak reading* of donkey sentences (Schubert and Pelletier, 1987).

In (149), given the quantificational force of the adverb *usually*, we would expect the truth of the sentence to require the majority of farmer-donkey pairs to be such that the farmer in them is rich. However, in a scenario where nine poor farmers own one donkey each and one rich farmer owns ninety donkeys, this is true – due to the rich farmer participating in ninety such pairs – but we judge the sentence as false. Rather than quantifying over farmer-donkey pairs – whether in the guise of assignments or minimal situations –, we seem to be quantifying only farmers here. This unexpected behaviour of the proportional reading is known as the *proportion problem* (Kadmon, 1987).

While dynamic semantics has formulated a standard response to these kinds of problems – selective quantification –, Elbourne's implementation of D-type theory does not extend to the empirical data in (147) and (149). However, the dynamic response can be reconstructed in D-type theory to a degree. Here, I will outline one potential implementation, based on my proposal in Walker (2014), with a particular focus on the proportion problem.

### 3.2.1 Selective quantification in dynamic semantics

The dynamic semantics solution to donkey sentences, in the standard DPL version sketched in Ch. 2, has the quantificational adverb – or, alternatively, the conditional itself – quantify over complete assignments, that is, if the antecedent introduces two variables – a farmer and a donkey –, then the quantification is over both their values. In effect, this amounts to quantification over farmer-donkey pairs. This is usually refered to as *unselective quantification* (Groenendijk and Stokhof, 1991; Lewis, 1975). In order to account for proportional readings, we introduce the possibility of quantifying only over a subset of the variables introduced by the antecedent, consequently known as *selective quantification* (Brasoveanu, 2013; Dekker,

1993; Root, 1986; van Rooij, 2006). As Dekker points out, the indefinite that is not quantified over still needs to modify assignments in the usual way, in order to realize its anaphoric potential, so that the solution must be placed in the quantificational adverb.

In DPL, the lexical entry for any (unselective) quantificational adverb is simply a generalization of the conditional (which is taken to have universal quantificational force), repeated below as (150) and re-written in set notation in (151), with the quantificational force Q – a generalized quantifier in the sense of Barwise and Cooper (1981) – replacing the universal quantifier – see (152):

(150) $[\![\phi \rightarrow \psi]\!] = \{\langle g, h\rangle \mid h = g \wedge \forall k : \langle h, k\rangle \in [\![\phi]\!] \rightarrow \exists j : \langle k, j\rangle \in [\![\psi]\!]\}$

(151) $[\![\phi \rightarrow \psi]\!] = \{\langle g, h\rangle \mid h = g \wedge \{k \mid \langle h, k\rangle \in [\![\phi]\!]\} \subseteq \{k \mid \exists j : \langle k, j\rangle \in [\![\psi]\!]\}\}$

(152) $[\![\phi \rightarrow_Q \psi]\!] = \{\langle g, h\rangle \mid h = g \wedge Q(\{k \mid \langle h, k\rangle \in [\![\phi]\!]\}, \{k \mid \exists j : \langle k, j\rangle \in [\![\psi]\!]\})\}$

We can see in (151) that the conditional in DPL expresses that the subset relation holds between the antecedent-satisfying assignments (relative to $h$) and the assignments that can serve as input for the consequent, i.e. its domain (Brasoveanu, 2013). Consequently, the respective static determiners Q relate the same sets with their corresponding quantificational force in (152).

In order to account for selective quantification, we need the static determiner Q to relate different sets. How exactly these are spelled out differs in the literature, but they all amount to the same idea: Instead of quantification over farmer-donkey pairs, we simply want to achieve quantification over farmers. Dekker (1993), e.g., implements this by quantifying over equivalence classes of assignments (those that only differ in what donkey they assign), while Brasoveanu (2013) suggests a formalization directly in terms of individuals. Both result in the same – here, we will follow Brasoveanu, to avoid introducing Dekker's additional (and unrelated) innovations.

In defining selective quantification, we will make use of some abbreviations defined in Brasoveanu (2013): $g[x]h$ expresses the random pointwise modification of $g$ in the variable $x$, i.e that $h$ differs from $g$ at most in the value of $x$. In a similar vein, we can treat $[x]$ as a formula in itself, where $[\![[x]]\!]$ is defined as the set of tuples $\langle g, h\rangle$ such that $g[x]h$, see (153). Brasoveanu then uses the abbreviation in (154) to construct a set of individuals for quantification:

(153) $[\![[x]]\!] = \{\langle g, h\rangle \mid h[x]g\}$

(154) $\{x \mid g[\![\phi]\!]\} := \{h(x) : \langle g, h\rangle \in [\![[x] \wedge \phi]\!]\}$

That is, (154) defines a set of individuals relative to an input assignment $g$ and a formula $\phi$ that is obtained by collecting those individuals assigned to $x$ in $h$ such that $h$ can be obtained by updating $g$ with the conjunction of the pointwise manipulation of $x$ and the formula $\phi$.

Less abstract, we can use this in order to extract the farmers $x$ from our antecedent $\phi$: We will get all those individuals that can be assigned to $x$ while the resulting assignments still verify the antecedent. These individuals are then what selective quantification quantifies over. However, we also need to construct the second set of individuals that the static quantifier relates these individuals to. We obtain this set by extracting the individuals from an update with both the antecedent and the consequent. However, these two updates can be related in two different ways: Conjoining them with dynamic conjunction will give us a weak, or asymmetric reading – see (155) –, while connecting them with a dynamic conditional will return the previous strong or symmetric reading – see (156). Note that this definition is also relative to a variable $x$, from which individuals are extracted. For further discussion of this result and its systematic relationships with the nature of DPL, as well as later developments in the system, see Brasoveanu (2013). Here, we will content ourselves with this preliminary result and its potential to be reconstructed in D-type theory.

(155) $[\![\phi \rightarrow_{Q-x-weak} \psi]\!] = \{\langle g, h\rangle \mid h = g \wedge Q(\{x \mid h[\![\phi]\!]\}, \{x \mid h[\![\phi \wedge \psi]\!]\})\}$

(156) $[\![\phi \rightarrow_{Q-x-strong} \psi]\!] = \{\langle g, h\rangle \mid h = g \wedge Q(\{x \mid h[\![\phi]\!]\}, \{x \mid h[\![\phi \rightarrow \psi]\!]\})\}$

As the strong version comes out equivalent to unselective quantification, I will only provide a computation for the weak version in (155). Consider the following sentence and formula:

(157) If a farmer owns a donkey, he beats it.

(158) $\exists x(Fx \wedge \exists y(Dy \wedge Oxy)) \rightarrow_{Q-x-weak} Bxy$

(158) is a slightly more complicated variant of our formula in (41b), which we have shown to be equivalent to $\forall x(Px \rightarrow Qx)$[4]. There are two main differences, motivated by our need to demonstrate weak readings: It contains two variables – $x$ providing the set of individuals to be quantified over, and $y$ being treated weakly –, and we are employing the weak conditional from (155). If we compute this variant, we obtain the result in (164).

For easier reference though, let me first repeat the standard DPL definitions from (36)–(40), which correspond to (159)–(163) respectively:

(159) $[\![Rt_1...t_n]\!] = \{\langle g,h\rangle \mid h = g \wedge \langle [\![t_1]\!]^h ... [\![t_n]\!]^h\rangle \in F(R)\}$

(160) $[\![\exists x\phi]\!] = \{\langle g,h\rangle \mid \exists k : k[x]g \wedge \langle k,h\rangle \in [\![\phi]\!]\}$

(161) $[\![\phi \wedge \psi]\!] = \{\langle g,h\rangle \mid \exists k : \langle g,k\rangle \in [\![\phi]\!] \wedge \langle k,h\rangle \in [\![\psi]\!]\}$

(162) $[\![\phi \rightarrow \psi]\!] = \{\langle g,h\rangle \mid h = g \wedge \forall k : \langle h,k\rangle \in [\![\phi]\!] \rightarrow \exists j : \langle k,j\rangle \in [\![\psi]\!]\}$

(163) $[\![\forall x\phi]\!] = \{\langle g,h\rangle \mid h = g \wedge \forall k : k[x]g \rightarrow \exists j : \langle k,j\rangle \in [\![\phi]\!]\}$

(164) a. $[\![\exists y(Dy \wedge Oxy)]\!]$
$= \{\langle g,h\rangle \mid h[y]g \wedge h(y) \in F(D)$
$\wedge \langle h(x),h(y)\rangle \in F(O)\}$ by (159), (160), (161)

b. $[\![Fx \wedge \exists y(Dy \wedge Oxy)]\!]$
$= \{\langle g,h\rangle \mid \exists k : \langle g,k\rangle \in [\![Fx]\!]$
$\wedge \langle k,h\rangle \in [\![\exists y(Dy \wedge Oxy)]\!]\}$ by (161)
$= \{\langle g,h\rangle \mid \exists k : g = k \wedge g(x) \in F(F)$
$\wedge \langle k,h\rangle \in [\![\exists y(Dy \wedge Oxy)]\!]\}$ by (159)
$= \{\langle g,h\rangle \mid \exists k : g = k \wedge g(x) \in F(F)$
$\wedge h[y]k \wedge h(y) \in F(D) \wedge \langle h(x),h(y)\rangle \in F(O)\}$ by (164a)

[4] For simplicity, we will be returning to classical DPL without possible worlds here.

$= \{\langle g, h\rangle \mid g(x) \in F(F)$
$\wedge\, h[y]g \wedge h(y) \in F(D) \wedge \langle h(x), h(y)\rangle \in F(O)\}$ by $g = k$

c. $[\![\exists x(Fx \wedge \exists y(Dy \wedge Oxy))]\!]$
$= \{\langle g, h\rangle \mid \exists k : k[x]g$
$\wedge\, \langle k, h\rangle \in [\![Fx \wedge \exists y(Dy \wedge Oxy)]\!]\}$ by (160)
$= \{\langle g, h\rangle \mid \exists k : k[x]g$
$\wedge\, k(x) \in F(F) \wedge h[y]k \wedge h(y) \in F(D)$
$\wedge\, \langle h(x), h(y)\rangle \in F(O)\}$ by (164b)

d. $[\![\exists x(Fx \wedge \exists y(Dy \wedge Oxy)) \wedge Bxy]\!]$
$= \{\langle g, h\rangle \mid \exists j : \langle g, j\rangle \in [\![\exists x(Fx \wedge \exists y(Dy \wedge Oxy))]\!]$
$\wedge\, \langle j, h\rangle \in [\![Bxy]\!]\}$ by (161)
$= \{\langle g, h\rangle \mid \exists j : \exists k(k[x]g \wedge k(x) \in F(F) \wedge j[y]k$
$\wedge\, j(y) \in F(D) \wedge \langle j(x), j(y)\rangle \in F(O))$
$\wedge\, \langle j, h\rangle \in [\![Bxy]\!]\}$ by (164c)
$= \{\langle g, h\rangle \mid \exists j : \exists k(k[x]g \wedge k(x) \in F(F) \wedge j[y]k$
$\wedge\, j(y) \in F(D) \wedge \langle j(x), j(y)\rangle \in F(O))$
$\wedge\, j = h \wedge \langle j(x), j(y)\rangle \in F(B)\}$ by (159)
$= \{\langle g, h\rangle \mid \exists k(k[x]g \wedge k(x) \in F(F) \wedge h[y]k$
$\wedge\, h(y) \in F(D) \wedge \langle h(x), h(y)\rangle \in F(O))$
$\wedge\, \langle h(x), h(y)\rangle \in F(B)\}$ by $j = h$

e. $[\![[x] \wedge \exists x(Fx \wedge \exists y(Dy \wedge Oxy))]\!]$
$= \{\langle g, h\rangle \mid \exists j : \langle g, j\rangle \in [\![[x]]\!]$
$\wedge\, \langle j, h\rangle \in [\![\exists x(Fx \wedge \exists y(Dy \wedge Oxy))]\!]\}$ by (159)
$= \{\langle g, h\rangle \mid \exists j : j[x]g$
$\wedge\, \langle j, h\rangle \in [\![\exists x(Fx \wedge \exists y(Dy \wedge Oxy))]\!]\}$ by (153)
$= \{\langle g, h\rangle \mid \exists j : j[x]g$
$\wedge\, \exists k(k[x]j \wedge k(x) \in F(F)$
$\wedge\, h[y]k \wedge h(y) \in F(D) \wedge \langle h(x), h(y)\rangle \in F(O)\}$ by (164c)

$= \{\langle g, h\rangle \mid \exists k(k[x]g \wedge k(x) \in F(F) \wedge k(x)$
$\wedge\, h[y]k \wedge h(y) \in F(D) \wedge \langle h(x), h(y)\rangle \in F(O)\}$ by $j[x]g \wedge k[x]j$ [5]

f. $[\![[x] \wedge (\exists x(Fx \wedge \exists y(Dy \wedge Oxy)) \wedge Bxy)]\!]$
$= \{\langle g, h\rangle \mid \exists j : j[x]g$
$\wedge\, \langle j, h\rangle \in [\![\exists x(Fx \wedge \exists y(Dy \wedge Oxy)) \wedge Bxy]\!]\}$ by (36), (153)
$= \{\langle g, h\rangle \mid \exists j : j[x]g$
$\wedge\, \exists k(k[x]j \wedge k(x) \in F(F)$
$\wedge\, h[y]k \wedge h(y) \in F(D) \wedge \langle h(x), h(y)\rangle \in F(O)$
$\wedge\, \langle h(x), h(y)\rangle \in F(B)\}$ by (164d)
$= \{\langle g, h\rangle \mid \exists k(k[x]g \wedge k(x) \in F(F)$
$\wedge\, h[y]k \wedge h(y) \in F(D) \wedge \langle h(x), h(y)\rangle \in F(O)$
$\wedge\, \langle h(x), h(y)\rangle \in F(B)\}$ by $j[x]g \wedge k[x]j$

g. $[\![\exists x(Fx \wedge \exists y(Dy \wedge Oxy)) \rightarrow_{Q-x-weak} Bxy]\!]$
$= \{\langle g, h\rangle \mid h = g \wedge$
$Q(\{x \mid g[\![\exists x(Fx \wedge \exists y(Dy \wedge Oxy))]\!]\},$
$\{x \mid g[\![\exists x(Fx \wedge \exists y(Dy \wedge Oxy)) \wedge Bxy]\!]\})\}$ by (155)
$= \{\langle g, h\rangle \mid h = g \wedge$
$Q(\{j(x) \mid \exists k(k[x]g \wedge k(x) \in F(F)$
$\wedge\, j[y]k \wedge j(y) \in F(D) \wedge \langle j(x), j(y)\rangle \in F(O)\},$
$\{x \mid g[\![\exists x(Fx \wedge \exists y(Dy \wedge Oxy)) \wedge Bxy]\!]\})\}$ by (154), (164e)
$= \{\langle g, h\rangle \mid h = g \wedge$
$Q(\{j(x) \mid \exists k(k[x]g \wedge k(x) \in F(F)$
$\wedge\, j[y]k \wedge j(y) \in F(D) \wedge \langle j(x), j(y)\rangle \in F(O)\},$
$\{j(x) \mid \exists k : k[x]g \wedge k(x) \in F(F)$
$\wedge\, h[y]k \wedge h(y) \in F(D) \wedge \langle h(x), h(y)\rangle \in F(O)$
$\wedge\, \langle h(x), h(y)\rangle \in F(B)\})\}$ by (154), (164f)

[5] Since we are placing no further conditions on $j$, applying pointwise manipulation of $x$ twice does not make a difference, so that we can remove $j$ and directly require $k[x]g$. Note that this renders (164e) and (164c) equivalent. The same holds for (164d) and (164f).

As you can see in (164g), the weak conditional relates two sets of individuals with the specified (static) quantifier Q: First, a set of individuals that satisfy the antecedent (i.e. farmers owning a donkey), and second, a set of individuals satisfying the antecedent and the consequent (i.e. farmers owning and beating a donkey). This yields the low reading, as a farmer will be included in the second set even if he only beats one of his donkeys. In contrast, a strong reading, employing the dynamic conditional rather than dynamic conjunction for constructing the second set, would only include individuals who beat all of their donkeys, yielding the familiar strong reading that is equivalent to unselective quantification.

### 3.2.2 A D-type theory solution?

Having described the dynamic solution, we will now consider a D-type variant, based on the suggestions in Walker (2014). Standard D-type theory (Elbourne, 2005) runs into exactly the same proportion problem as dynamic semantics, as it quantifies over minimal situations containing a unique farmer and a unique donkey, i.e., over farmer-donkey pairs. In the spirit of the dynamic solution, we are first going to reconstruct Elbourne's quantificational adverb *always* as relating two sets of situations, and then extend it to generalized quantifiers[6].

Elbourne's *always* is repeated below from (140), and re-written in set notation in (166). Here, we have the subset relation holding between the set of minimal situations satisfying the antecedent, and those minimal situations satisfying the antecedent for which there is a minimal extension that satisfies the consequent. This allows us to employ the standard generalized quantifiers, but deviates from my original proposal in Walker (2014) where *always* directly relates the antecedent-situations with their extensions by requiring equal cardinality of the sets.

[6] Elbourne's original proposal does not contain quantificational adverbs beyond *always*, but I think it is reasonable to assume the extension to generalized quantifiers proposed here would be roughly in his spirit.

(165) $[\![always]\!]^g = \lambda p_{\langle s,t\rangle}.\lambda q_{\langle s,t\rangle}.\lambda s.$ for every minimal situation $s'$ such that $s' \sqsubseteq s$ and $p(s') = 1$, there is a situation $s''$ such that $s'' \sqsubseteq s$ and $s''$ is a minimal situation such that $s' \sqsubseteq s''$ and $q(s'') = 1$

(166) $[\![always]\!]^g = \lambda p_{\langle s,t\rangle}.\lambda q_{\langle s,t\rangle}.\lambda s.\{s' \mid s'$ is a minimal situation such that $s' \sqsubseteq s \wedge p(s') = 1\} \subseteq \{s' \mid \exists s'' : s'$ is a minimal situation such that $s' \sqsubseteq s \wedge p(s') = 1 \wedge s''$ is a minimal situation such that $s'' \sqsubseteq s \wedge s' \sqsubseteq s'' \wedge q(s'') = 1\}$

We can then give a general lexical entry for generalized quantifiers, as in (167), exemplified for *usually* in (168), given the simplifying assumption that *usually / most* expresses "more than half":

(167) $[\![Q]\!]^g = \lambda p_{\langle s,t\rangle}.\lambda q_{\langle s,t\rangle}.\lambda s.Q(\{s' \mid s'$ is a minimal situation such that $s' \sqsubseteq s \wedge p(s') = 1\}, \{s' \mid \exists s'' : s'$ is a minimal situation such that $s' \sqsubseteq s \wedge p(s') = 1 \wedge s''$ is a minimal situation such that $s'' \sqsubseteq s \wedge s' \sqsubseteq s'' \wedge q(s'') = 1\})$

(168) $[\![usually]\!]^g = \lambda p_{\langle s,t\rangle}.\lambda q_{\langle s,t\rangle}.\lambda s.\ \frac{1}{2}\ |\{s' \mid s'$ is a minimal situation such that $s' \sqsubseteq s \wedge p(s') = 1\}| < |\{s' \mid \exists s'' : s'$ is a minimal situation such that $s' \sqsubseteq s \wedge p(s') = 1 \wedge s''$ is a minimal situation such that $s'' \sqsubseteq s \wedge s' \sqsubseteq s'' \wedge q(s'') = 1\}|$

(168) expresses that the number of consequent-extendable antecedent situations is greater than one half of the number of all antecedent situations. For simplicity, we will now introduce the following abbreviations to refer to these two sets of situations:

(169) ANT := $\{s' \mid s'$ is a minimal situation such that $s' \sqsubseteq s \wedge p(s') = 1\}$

(170) ANT+ := $\{s' \mid \exists s'' : s'$ is a minimal situation such that $s' \sqsubseteq s \wedge p(s') = 1 \wedge s''$ is a minimal situation such that $s'' \sqsubseteq s \wedge s' \sqsubseteq s'' \wedge q(s'') = 1\}$

ANT refers to the antecedent-situations (labelled $s_4$ in (145)), ANT+ refers to those antecedent-situations which have a consequent-verifying extension (labelled $s_5$ in (145)). Given these abbreviations, we can rewrite (167) and (168) as follows:

(171) $[\![Q]\!]^g = \lambda p_{\langle s,t\rangle}.\lambda q_{\langle s,t\rangle}.\lambda s.Q(\text{ANT}, \text{ANT+})$

(172) $[\![usually]\!]^g = \lambda p_{\langle s,t\rangle}.\lambda q_{\langle s,t\rangle}.\lambda s.\ \frac{1}{2}\ |\textsc{Ant}| < |\textsc{Ant+}|$

(172) will give us the standard quantification over farmer-donkey pairs that runs into the proportion problem. In Walker (2014) I discuss two potential strategies for avoiding this problem: Either, one could adapt the semantics of the indefinites that constitute the sets Ant and Ant+, or one can implement a solution directly in the semantics of the quantificational adverb. The first strategy, however, immediately runs into trouble, because we employ the situations in Ant and Ant+ for more than one purpose: They are relevant to our quantification here, but they also guarantee the anaphoric potential of the indefinite through their minimality requirement, providing a unique referent for the definite description in the consequent.

To see the problem, have a look at our definition of Ant+ in (170): We are looking for situations that both verify the consequent $q$ and are extensions of situations that verify the antecedent $p$. At the moment, situations that verify $p$ will always be minimal situations that contain at most one farmer and one donkey, due to the lexical entry of the indefinite specified in (138). If we were to change this lexical entry in such a way that it would, e.g., return situations of farmers together with all their donkeys, in order to solve the proportion problem, our extensions of these situations would also contain multiple donkeys, making it impossible for the definite description to refer to the unique donkey in that situation. In effect, we would have re-introduced the problem of uniqueness presuppositions that situation semantics was supposed to solve into D-type theory.

Like in dynamic semantics, it therefore seems more promising to work directly on the quantificational adverb instead (but see Brasoveanu (2008) and Brasoveanu and Dotlacil (2016) for the challenge that mixed weak / strong readings pose to the dynamic implementation sketched above). In order to preserve the anaphoricity of indefinites, we will leave Ant and Ant+ – or rather, the indefinite that determines what situations verify $p$ – untouched. However, the semantics of the quantificational adverb will not let the generalized quantifier directly relate these two sets. Rather, they will

both be further processed first, yielding the actual sets to be compared. The processing will be done by an operation that I call Fuse in Walker (2014). Fuse applies to both Ant and Ant+, so that we obtain a lexical entry like (173).

(173) $[\![usually]\!]^g = \lambda p_{\langle s,t\rangle}.\lambda q_{\langle s,t\rangle}.\lambda s.\ \frac{1}{2}\ |\text{Fuse}(\text{Ant})| < |\text{Fuse}(\text{Ant+})|$

Now, since we are simply dealing with a generalized quantifier relating two sets again, Fuse will have to make it possible for us to obtain the same effect as in dynamic semantics: We want to end up quantifying over individuals, either directly or indirectly (as, e.g., in Dekker's (1993) account using equivalence classes of assignments). That is, Fuse needs to either yield sets of individuals, or sets of entities that can suitably stand in for those individuals. Either way, these entities need to be systematically derivable from the minimal situations describing the antecedent and the consequent.

In dynamic semantics, we can simply extract individuals from assignments by specifying a variable, as shown in 3.2.1. In a D-type theory framework, we do not have this possibility: Once the nested situation structure is generated, there is no record of the fact that all farmers in the minimal farmer-donkey situations are generated by the same indefinite. This creates a challenge for deriving a list of individuals (or individual-like entities) from minimal situations.

In Walker (2014), I propose the following semantics for Fuse[7]:

(174) $\text{Fuse}_K = \lambda S_{\langle s,t\rangle}.\{$ the minimal situation $s$ such that
$\forall s' \in S : x \sqsubseteq s' \rightarrow s' \sqsubseteq s \mid x \in K\}$
where $K$ is a contextually given set of individuals

[7] In Walker (2014), I first propose this semantics and then revise it to cover cases with an arbitrary number of variables to account for mixed readings. However, as my following criticism applies to both the simple version presented here and the revised version from the paper, I limit the present discussion to (174) for reasons of simplicity. The proposed change assumes a contextual variable consisting of n-tuples of individuals; for details, the reader is referred to the discussion in the paper.

That is, FUSE is a function from sets of situations to sets of situations. It takes the original sets ANT and ANT+ in $S$ and returns sets that contain those minimal situations which can be generated from the following algorithm: For each subset of situations from $S$ that contain a given individual $x$ from the contextual variable $K$, they return one minimal situation that contains just those situations as parts. If $K$ contains all farmers, and $S$ is a set of minimal farmer-donkey situations, then FUSE returns one minimal situation per farmer, containing just that farmer, all of his donkeys and the owning relationships that hold between the farmer and his donkeys.

Consider the following model in (175) and sentence in (176) with the associated sets ANT and ANT+in (177) and (178). I use square brackets [...] to abbreviate "the minimal situation such that ..." – e.g., [$f_1$ owns $d_1$] should be read as "the minimal situation such that $f_1$ owns $d_1$."

(175) There are three farmers. Farmer 1 ($f_1$) and Farmer 2 ($f_2$) have one donkey each ($d_1, d_2$) which they do not beat, but Farmer 3 ($f_3$) beats all of his three donkeys ($d_3, d_4, d_5$).

(176) If a farmer owns a donkey, he usually beats it.

(177) ANT = { [$f_1$ owns $d_1$], [$f_2$ owns $d_2$], [$f_3$ owns $d_3$], [$f_3$ owns $d_4$], [$f_3$ owns $d_5$] }

(178) ANT+ = { [$f_3$ owns $d_3$], [$f_3$ owns $d_4$], [$f_3$ owns $d_5$] }

Comparing proportions without applying FUSE will lead to the undesired standard result of (176) being true: A majority ($\frac{3}{5}$) of situations in ANT also appears in ANT+. However, applying FUSE – with $K = \{f_1, f_2, f_3\}$ – to both sets yields the following result:

(179) FUSE(ANT) = { [$f_1$ owns $d_1$], [$f_2$ owns $d_2$], [$f_3$ owns $d_3, d_4$ and $d_5$] }

(180) FUSE(ANT+) = { [$f_3$ owns $d_3, d_4$ and $d_5$] }

Since all situations involving $f_3$ have been fused together, now only a minority ($\frac{1}{3}$) of situations in FUSE(ANT) also appears in FUSE(ANT+).

Note that, differently from selective quantification, this cannot account for weak readings like (147) at the same time as it does for proportional readings. Because we are not truly quantifying over individuals, but over situations consisting of individuals and some of their dependents (like donkeys or dimes), the subset relation of *always* will not hold between a fused set of men and their dimes and a fused set of men and their thrown dimes if there are men who do not throw all of their dimes. In Walker (2014), I escape this result by using a cardinality-based definition for *always* that relates antecedent-situations with fully extended consequent-situations, allowing me to generate asymmetric and weak readings with the same mechanism, as suggested by dynamic semantics[8]:

(181) $[\![always]\!]^g = \lambda p_{\langle s,t\rangle}.\lambda q_{\langle s,t\rangle}.\lambda s.|\textsc{Fuse}(\textsc{Ant})| = |\textsc{Fuse}(\textsc{Cons})|$

(182) $\textsc{Cons} := \{s'' \mid \exists s' : s'$ is a minimal situation such that $s' \sqsubseteq s \wedge p(s') = 1 \wedge s''$ is a minimal situation such that $s'' \sqsubseteq s \wedge s' \sqsubseteq s'' \wedge q(s'') = 1\}$

The difference between Ant+ and Cons is the following: Ant+, as defined in (170), consists of situations $s'$ for which there exists an extension $s''$ that verifies the consequent. In contrast, Cons, as defined in (182), consists of exactly those extensions $s''$. If one minimal situation $s'$ has multiple possible extensions, Ant+ will contain only that potentially extendable situation, while Cons will contain multiple situations, namely all the possible extensions. Since these situations are different from the original non-extended situations, we cannot successfully apply the subset relation when using Cons. Instead, in Walker (2014), I apply the cardinality-based quantifier in (181). Consider the following example:

(183) If a man has a dime, he throws it into the meter.

(184) There are two men with two dimes each, who each throw one of their dimes into the meter.

[8] Note that this result does not strictly depend on the choice between Ant+ and Cons, but rather on the cardinality-based quantifier. Below, I will follow Walker (2014) in employing Cons, as defined in (182), but the results carry over to Ant+ as long as (181) is used. However, the non-cardinality-based approach cannot employ Cons, because it depends on the identity between situations in Ant and Ant+ in order to establish subset relationships.

(185) Ant= { [$m_1$ has $d_{1a}$], [$m_1$ has $d_{1b}$], [$m_2$ has $d_{2a}$], [$m_1$ has $d_{2b}$] }

(186) Cons = { [$m_1$ has and throws $d_{1a}$], [$m_2$ has and throws $d_{2a}$] }

Under a strong (un-Fuseed) reading, this would come out as false, as there are four situations in Ant, but only two in Cons. However, applying Fuse will now yield the weak reading:

(187) Fuse(Ant) = { [$m_1$ has $d_{1a}$ and $d_{1b}$], [$m_2$ has $d_{2a}$ and $d_{2b}$] }

(188) Fuse(Cons) = { [$m_1$ has and throws $d_{1a}$], [$m_2$ has and throws $d_{2a}$] }

However, this departure from standard generalized quantifiers in favour of a cardinality-based version throws doubt on our ability to generalize these results and leaves room for accidental overlap in cardinality without any principled relationship between the two sets. That is, we can either use the non-cardinality-based version from this chapter and give up on a unified account for weak and proportional readings, or we can use the cardinality-based version from Walker (2014) and potentially introduce problems for our theory of generalized quantifiers.

Let's take a step back. Is there a way we could revise D-type theory in a more general way, avoiding both of these problems? One possibility would be to somehow make the content of sub-situations below the level of the antecedent and consequent accessible to the interpretation at a later point by attaching a label to them in the meta-language – e.g. following the numbering scheme in (145), to call the set of farmers $S_7$, which would allow us to extract a set of individuals in the same way dynamic semantics does in (154). We could then imagine a solution that avoids the complexities of Fuse and its interactions with situations by simply referring to these labels. But trouble abounds: First, we would need to then ensure that the individuals selected by the definite descriptions in the consequent are also labeled, second, that the two labels would somehow be linked with one another – in short, we would have to recreate a system that looks exactly like the indices on variables in dynamic semantics[9] .

[9] Something along these lines is also considered as a solution to the bishop problem, e.g., in Heim (1990). See also Brasoveanu and Dotlacil (2016), who share my worry that this simply recreates dynamic semantics.

As it stands, we can see that D-type theory can account for proportional readings, however, in a form that is much more limited than the dynamic counterpart, and which, in attempting to fix the remaining issues, very much begins to resemble a dynamic system. The main issue is that D-type theory does not hold on to the same amount of information as dynamic semantics – while we are essentially operating on farmer-donkey pairs in both of them, only dynamic semantics retains the information which of the two has been introduced by which indefinite. This is particularly problematic when we consider the following variant of the problem of indistinguishable participants:

(189) If a farmer hires a farmhand, he usually pays him well.

While (189) does not run into the bishop problem – there is a unique farmer and farmhand respectively, clearly identifiable via the descriptive content –, it runs into a problem when fusing situations based on contextually supplied lists of individuals, rather than abstract labels like variable names. Imagine a situation in which there is a farmer who for some reason works as a farmhand in his spare time – maybe he has enough hired help on his own farm, or he is in need of money, or both. Asymmetric quantification employing Fuse will now merge situations in which he appears as the farmer with those in which he appears as the farmhand, creating unpredictable, and unwanted results. Consider the following model:

(190) There are three farmers, $f_1$ to $f_3$, who all hire a fourth farmer, Gaston, as their farmhand and pay him badly. Since this leaves Gaston no time to run his own farm, he hires LeFou – who is not a farmer – as his farmhand and pays him well. The last farmer, Maurice, hires Belle – also not a farmer – and pays her well, too.

(191) Ant = { [$f_1$ hires Gaston], [$f_2$ hires Gaston], [$f_3$ hires Gaston], [Gaston hires LeFou], [Maurice hires Belle] }

(192) Ant+ = { [Gaston hires LeFou], [Maurice hires Belle] }

Given this setup, we would expect (189) to come out as false – after all, a majority of farmers pays their farmhands badly, with the exception of Gaston and Maurice. And so there are only two situations in ANT+ but five in ANT. But now we apply FUSE with $K$ containing all farmers, i.e. {$f_1$, $f_2$, $f_3$, Gaston, Maurice}:

(193) FUSE(ANT) = { [$f_1$ hires Gaston, $f_2$ hires Gaston, $f_3$ hires Gaston and Gaston hires LeFou], [Maurice hires Belle] }

(194) FUSE(ANT+) = { [Gaston hires LeFou], [Maurice hires Belle] }

Because FUSE cannot distinguish between participant roles in situations, all situations that involve Gaston in ANT get merged into one large situation. But now the cardinality of both FUSE(ANT) and FUSE(ANT+) is equally 2, so that (189) actually comes out as true: $\frac{1}{2}$ of 2 is indeed smaller than 2.

This result has interesting theoretical repercussions: It demonstrates that even in a D-type framework, we still need some form of co-indexation between indefinites and the definites that refer back to them in order to provide a full solution to both the proportion problem and weak readings. This shows that the two frameworks are much more closely related than they are often considered to be, as soon as both are extended to cover the full range of data, a point also made by Brasoveanu and Dotlacil (2016) in their overview of donkey theories. In many cases, the right question to ask is not "D-type theory or dynamic semantics?" but rather "which ingredients from which theory?"

## 3.3 Counterfactuals

### 3.3.1 Elbourne's account

Elbourne (2013) extends standard D-type theory to counterfactual conditionals. He proposes the following semantics for the counterfactual, which he treats in parallel to the quantificational adverbs of indicative conditionals (the following lexical entry makes use of the abbreviation $w_s$, denoting the (unique) world $w$ that contains $s$):

(195) $[\![CF]\!]^g = \lambda p_{\langle s,t\rangle}.\lambda q_{\langle s,t\rangle}.\lambda s.$ for every minimal situation $s'$ such that $p(s') = 1$ and $w'_s$ is otherwise as similar as possible to $w_s$, there is a situation $s''$ such that $s''$ is a minimal situation such that $s' \sqsubseteq s''$ and $q(s'') = 1$

As indicated by Elbourne, and evident from the proposed semantics, this is a straightforward implementation of Lewis-Stalnaker's variably strict semantics. It ranges over minimal antecedent-situations in all antecedent-worlds that are maximally similar to the actual world. Consequently, it also predicts only the low reading, as discussed in (2.2.2).

### 3.3.2 High and low readings: a first attempt

In what follows, we will again attempt to solve the puzzle of high and low readings along the lines of the solution suggested by dynamic semantics. That is, we will work off the assumption that the solution mainly consists in a partialization of similarity on the basis of individuals.

The dynamic solution discussed in Ch. 2 implements this partialization through assignments. As in our account of the proportion problem, we lack this tool and consequently dynamic semantic's ability to keep track of how individuals have been introduced into the discourse. In 3.2.2, we made an attempt to recreate the dynamic mechanics through other means. Here, we will pursue a similar strategy, demonstrating in course the challenges it has to meet.

Where van Rooij (2006) quantifies over world-assignment pairs, Elbourne's semantics quantifies over situations. In the same way that dynamic semantics employed a similarity relation that did not render all world-assignment pairs comparable, we want to obtain a similarity relation that does not render all situations comparable. Note that in (195), we simply quantify over the situations within the overall closest world. Instead, we now want to have the optional capability to quantify over situations within the closest worlds for their respective participants verifying the antecedent, that is, e.g., for each owned donkey separately.

One potential solution builds on Arregui's notion of modal parthood (Arregui, 2009). While we will return to Arregui's full account in a later chapter, a brief explanation of this concept is in order here: In Kratzer's situation semantics framework, worlds are maximal situations, that is, situations that are not contained within any other situation. At the same time, situations are at most part of one such maximal situation. Like individuals, they are world-bound, that is, they exist in only one world at a time. For this reason, if we want to talk about individuals or situations across worlds, we need to identify them with one another via a counterpart relation (Kratzer, 1989; Lewis, 1968)[10]. There are numerous philosophical difficulties in deciding how exactly such a relation can be constituted (Lewis, 1968, 1971, 1973b, 1983), but for present purposes it will suffice to assume it to be a primitive, contextually given relation, very much like the similarity relation (and indeed possibly based on it). Given such a counterpart relation, which we will write as $\approx$, we can then define a notion of "parthood" that works across worlds, despite situations being world-bound (Arregui, 2009):

(196) $s_i \sqsubseteq_m s_j$ iff $\exists s_t : s_t \sqsubseteq s_j \wedge s_t \approx s_i$
and $\approx$ is a suitably defined counterpart relation between situations.

This gives us a way of (loosely) talking about the "same" situation being present in two different worlds: If a situation is a modal part of another world, then that world contains a counterpart of that situation, i.e. a situation that is "the same" up to whatever standard the similarity relation enforces. It is this flexibility in defining such standards that we can then subsequently exploit to come up with a semantics for high and low readings.

[10] Note that Elbourne (2013) gives up this assumption, instead adopting the position that thin individuals can appear in more than one possible world, in order to simplify his semantics for proper names. However, he still upholds the position that situations are world-bound. As this requires us to assume counterpart relations in our semantics anyhow, I see no direct benefit from giving them up for thin individuals alone, at least within the scope of the present discussion.

As situations in D-type theory stand in for individuals comparably to how assignments do in dynamic semantics, we can also use a similar way of letting situations partialize the similarity relation now. In order to do this, let us first take a somewhat closer look at Elbourne's proposed semantics in (195), repeated below as (197):

(197) $[\![CF]\!]^g = \lambda p_{\langle s,t\rangle}.\lambda q_{\langle s,t\rangle}.\lambda s.$ for every minimal situation $s'$ such that $p(s') = 1$ and $w'_s$ is otherwise as similar as possible to $w_s$, there is a situation $s''$ such that $s''$ is a minimal situation such that $s' \sqsubseteq s''$ and $q(s'') = 1$

Note that Elbourne's lexical entry restricts quantification to those antecedent-satisfying situations that are part of a world that is "otherwise as similar as possible to $w_s$". The set of worlds that Elbourne intends to target is the one returned by the standard Lewisian selection function $f$ for the antecedent $p$, i.e. (49) repeated below as (198):

(198) $f_w([\![p]\!]) = \{v \in [\![p]\!] \mid \neg\exists u \in [\![p]\!] : u <_w v\}$

We can make a slight adjustment here that changes nothing substantial but will help us refine this selection function in a second step. Instead of requiring $v$ to verify $p$, we simply require $v$ to contain a $p$-situation:

(199) $f_w([\![p]\!]) = \{v \mid \exists s : s \sqsubseteq v \wedge s \in [\![p]\!] \wedge \neg\exists u \in [\![p]\!] : u <_w v\}$

(198)/(199) returns those worlds $v$ that satisfy the antecedent for which there is no other world $u$ that also satisfies the antecedent and that is more similar to the actual world $w$. This yields the standard low reading, and van Rooij's (2006) proposed move in dynamic semantics is the replacement of the similarity ordering employed with one that is sensitive to assignments (that is, in essence, to individuals). In parallel, we are going to make our selection function sensitive to situations (and, indirectly, to the individuals participating in them).

We will weaken the requirement in (198) as follows in (200): A world $v$ will be returned by the selection function as long as there is no closer world $u$ that contains a counterpart of the situation verifying the antecedent in $v$. Through the inclusion of $\sqsubseteq_m$ in the selection function, we make

it sensitive to the contextually given counterpart relation $\approx$. We then need to define two cases: The low reading should be brought about by a counterpart relation which trivializes this additional requirement, while the high reading should have a counterpart relation that partitions the similarity relation in the same way as assignments do in the dynamic account.

(200) $f_w([\![p]\!]) = \{v \mid \exists s : s \sqsubseteq v \wedge s \in [\![p]\!] \wedge \neg\exists u : s \sqsubseteq_m u \wedge u <_w v\}$

There are two intuitively plausible counterpart relations that, put into the selection function in (200), will do exactly that. One is a counterpart relation that only looks at general structural features: For counterfactual donkey sentences, it considers all antecedent-verifying situations counterparts of one another based on the fact that they are all farmer-donkey-owning situations. It yields the low reading, as all antecedent-situations can be compared to one another, leading (200) to only return those situations in the absolutely closest world. On the other hand, we can generate the high reading by assuming a stricter counterpart relation that only treats two situations as counterparts of one another if they also agree in their participants. In that sense, a situation of John owning Platero would not be a counterpart of a situation in which John owns Grisella. Including the closest Platero-world has no effect on the inclusion of the closest Grisella-world, as they are not ranked with respect to one another. In effect, (200) returns the closest world for each individual separately. For mixed readings, we can assume counterpart relations that care about the identity of certain participants but not others respectively.

However, it is very clear that these particular counterpart relations are fairly stipulative, and that they seem to run counter to any motivating paradigm we might want to come up with: For example, if we wanted them to reflect the goals of the discourse participants, we would expect a scenario in which participants *do* care about the identity of a particular donkey – and thus aim for a low reading – to have a counterpart relation that has relatively strict standards regarding participants, whereas a scenario in which participants only want to speak about attitudes towards

donkeys in general – and thus aim for a high reading – should have a very permissive counterpart relation. However, the picture is reversed: The stricter counterpart relation generates the more general reading, and the more permissive one generates the reading that targets a particular donkey.

Of course, various stories might be concocted to justify this particular pattern, but in the absence of independent insights into the nature of counterpart relations, they will all remain *ad hoc* to a degree. We also cannot provide a precise formal account of how counterpart relations between situations target particular participant roles, potentially running us into either energetic handwaving or the same problems as we had in the case of the proportion problem.

Once again, we have – like dynamic semantics – succeeded in providing successful mechanics for this particular problem, but no further insights into the pragmatics of these readings are forthcoming on this route.

## 3.4 Summary and discussion

As both the proportion problem and our first stab at high and low readings have shown, there is little promise in reconstructing dynamic solutions in D-type theory. Invariably, we run into the same problems which, in order to be fixed, would require us to adopt the machinery from dynamic semantics that we were setting out to avoid in the first place. Nothing is gained by this strategy either, as we at best re-create a solution that had already been available to us.

Does this mean that we should abandon D-type theory and focus on dynamic semantics instead? Not necessarily. As Brasoveanu and Dotlacil (2016) point out, dynamic semantics in turn needs to incorporate many features that have originated in description-theoretic approaches. However, what these issues show is that we should not approach the problem

in an "arms race" mentality, attempting to solve the same problems over and over again in both frameworks. Rather, we should focus on what each framework most productively can contribute to an overall successful theory.

Given dynamic semantics' ability to keep track of discourse referents beyond their potential referents – especially of by which linguistic element they have been introduced –, it would seem advisable to employ this system to account for the indefinite's anaphoric potential, unless other theoretical considerations absolutely speak against employing a dynamic semantics. However, from employing these means in order to handle anaphora, it does not automatically follow that we should use the same tools for any problem arising from the use of an indefinite. As pointed out in the discussion of Ch. (2), indefinites are not necessarily an ingredient of counterfactual donkey sentences. This may allow us to instead pursue an analysis where the role of the indefinite for anaphoricity and its role for counterfactual phenomena come apart. We discuss such an analysis in the following Ch. 4.

# Chapter 4

# Similarity reconsidered

## 4.1 Towards an account of similarity

In the previous two chapters, we have discussed ways of integrating our observations on high and low readings with the two main lines of donkey theory, while keeping the standard Stalnaker-Lewis approach to counterfactuals fixed. In the dynamic line, we have followed van Rooij (2006) and used assignments to partialize the Lewisian similarity function. In D-type theory, we have tentatively reconstructed the dynamic solution, using counterpart relations between situations. However, as the discussion in both chapters has demonstrated, both solutions remain unsatisfactory.

In these discussions, I have already suggested where I see the problem with our previous attempts: In leaving the Lewisian similarity relation untouched, we treat high and low readings as an artefact of our semantics for the indefinite. But this cannot explain the radical dependence on context we observe for the distribution of both readings. It also does very little for scenarios of the "thieves" type in (202) discussed in Bennett (2003) and the subsequent philosophical literature (e.g. Khoo (2016), Nichols (2016), and Placek and Müller (2007)), or the "dart" example in (203) due to Schwarz (2013) unless we are comfortable postulating a large number of hidden indefinites:

(201) My coat was not stolen from the restaurant where I left it. There were two chances for theft [...]. They would have involved different potential thieves: and the candidate for the later theft is a rogue who always sells his stuff to a pawn-broker named Fence.

(202) If my coat had been stolen from the restaurant, it would now be in Fence's shop.

(203) If the dart had landed on the left side of the board, I would have won.

Despite having no indefinite in (202) we observe the same ambiguity between a high and a low reading, if we assume that one of the thieves is more likely than the other[1]. In the same way, (203), on its most natural (high) reading, entails that any location on the left side of the board would have been sufficient for a win, but again there is no indefinite or disjunction overtly present in the example.

A similar point is made by Arregui (2005b) with examples like the following:

(204) Jim and Jack danced very romantically all night. If Jim had danced with Joe, he would have danced very romantically too.

(205) Jim has spent the evening dancing with Jack by the fireplace. If Jim had danced with Joe, he would have danced by the fireplace (too).

As Arregui points out, (204) and (205) are not automatically true, despite the fact that dances that agree in the manner (romantically) or place (by the fireplace) should be the most similar ones. Apparently, a high reading over these arguments arises, even in the absence of an overt indefinite. I share Arregui's reluctance to conclude that this should lead us to posit covert indefinites for all such arguments, although I do not want to discard that possibility entirely.

[1] Bennett's example contains no such assumption about likelihood, because he is only attempting to demonstrate that the reading which corresponds to our high reading, and which we judge as false, arises. This serves as an argument against time-based approaches, where the latest divergence from the actual world is privileged, making (202) automatically true in the context of (201).

In this chapter, I am going to discuss another option: Reconsidering the classical Lewisian notion of similarity and replacing or amending it with something that gives us a better way of tracking context-dependence. In the discussion of proposals in the literature, I will limit myself to those aspects that directly concern counterfactual donkey sentences, but they often are motivated by other data that show an independent need for reconsidering the workings of similarity.

I will focus on three broad areas: causality, relevance and indexicality. While there is other literature critical of Lewis' notion of similarity – especially in philosophy, see e.g. Fine (1975), Tichý (1976), Slote (1978), Bennett (2003), Tooley (2003), Khoo (2016) –, it often aims to leave the Lewis-Stalnaker account behind entirely, or addresses issues that lie far outside the scope of this dissertation and our problem at hand. Within these three areas, however, there are proposals for upholding the general framework while improving our understanding of Lewis' notion of similarity in a way that sheds light on counterfactual donkey sentences, by adding causal structures and notions of relevance, or by proposing similarity to hold between situations rather than entire possible worlds, making it more tractable (and potentially including both causality and relevance).

After discussing the literature in these areas, I will proceed to my proposal. I will show how Arregui's (2009) proposal for an indexical similarity relation, which standardly generates high readings, can be made more precise by modelling the influence of laws through causal dynamics in the style of Schulz (2011). I will then demonstrate that the low reading arises from our ability to elevate regularities to law-like status.

## 4.2 Relevance

I am currently aware of two proposals for incorporating relevance into our analysis of counterfactuals. Common to both is that they are not aiming to wholly do away with similarity – they uphold its importance either explicitly (Lewis, 2016)[2] or implicitly (Nichols, 2016). But they attempt to account for various problems in Lewis's (1973) original conception by spelling out how similarity is mediated by relevance. Both accounts are motivated by slightly different problems. I will tackle Lewis's (2016) discussion of counterfactual skepticism first, and then proceed to Nichols's (2016), whose issues are more closely related to the problems discussed in this thesis.

### 4.2.1 Counterfactual skepticism

Counterfactual skepticism, the position against which Lewis (2016) develops her relevance account, is the claim that most or indeed all counterfactuals that we utter are false. Such a worry arises from different kinds of examples. Some are motivated by concerns about quantum physics, such as the following (DeRose, 1999):

(206) If you had dropped that vase, it would have broken.

(207) If you had dropped that vase, a quantum event might have happened in which it flew sideways and landed safely on the couch.

(208) If you had dropped that vase, it might not have broken.

[2] A note on the multiplicity of Lewises writing on miracles is in order – there are at least three. While C.S. Lewis does not make an appearance in this thesis, the other two do: Lewis (2016) is the work of Karen Lewis, not David Lewis.

The worry here is that we judge (206) as true in normal circumstances. However, given our current knowledge of quantum physics, (207) is also true, and since it entails (208) there is a clash between this judgment and our previous judgment for (206). I believe we can safely set aside this worry, as hearers rarely access modern quantum physics in their judgments. But as Lewis points out, the same problem arises in more normal scenarios as well[3]:

(209) If I had tagged up, I would have scored the winning run.

(210) If I had tagged up, I might have tripped, fallen, and been thrown out.

(211) If I had tagged up, I might not have scored the winning run.

Again, we are invited to observe the same clash.

Lewis' proposed solution in (212) is, in a way, related to von Fintel's (1999) proposal from Ch. 2:

(212) For all contexts c, $P \boxdot\!\!\rightarrow Q$ is true in c iff all the closest P-worlds are Q-worlds, where closeness is a function of both similarity and relevance.

The basic idea is that closeness – the relation that determines what worlds a counterfactual quantifies over – is not simply determined by similarity, but also by relevance. This allows the counterfactual to quantify over worlds that are not the most similar ones, as long as they are in some way relevant to the conversation. One explicit way of making a world relevant is via *might*-conditionals. Compare von Fintel's idea: We always quantify over the entire domain of worlds specified by the modal horizon. While similarity plays a crucial role in expanding the modal horizon, once a world is in the horizon (that is, has been made relevant to the conversation) it is always used for quantification, even if it is not amongst the most similar ones.

[3] More normal, that is, for someone familiar with baseball in a way that I am not. But I believe that the example works even without precisely understanding what's going on.

This allows Lewis to uphold the duality of *might* and *would* – without which, there would be no necessary clash between the claims in (209)–(211) –, while explaining the clashing judgments: Essentially, (209) is evaluated with respect to a different set of relevant worlds when used discourse-initially and when used after making other worlds relevant through uttering (210).

### 4.2.2 The specificity problem

Nichols (2016) comes to a similar proposal via different examples. One of them is the familiar problem of disjunctive antecedents which, as van Rooij (2006) points out, can be seen as a variant of the problem of counterfactual donkeys, given the interchangeability of indefinites and disjunctions:

(213) ? If I were in Miami or Havana, I'd be in the USA.

Note that Nichols (2016) unambiguously considers (213) false. For him, (213) comes out as false because it asserts both that if I were in Miami, I would be in the USA and that if I were in Havana, I would be in the USA, the latter obviously being false. I think this is not quite correct: In addition to this reading, there is an (admittedly, marginal) low reading on which the sentence comes out true. This reading is one in which I speculate, e.g., about the choices I could have made that would have put me in another place. I can then read (213) as expressing that given a choice between those two places, I would have chosen to be in the USA, that is, in Miami. Consider the following variant, which brings out this reading more clearly:

(214) If I were to go to Miami or Havana, I'd go to the USA.

As with indefinites, there is the possibility of obtaining low readings. This aside though, Nichols' point is familiar: The existence of a false reading of (213) is challenging for the standard Lewis-Stalnaker account. This is the basic problem of high readings.

Note that Nichols also points out that this does not seem to be explicitly tied to any lexical material, although it is most conveniently observed with indefinites and disjunctions. One counterexample is the case of the thief in (201)–(202) discussed in the introduction of this chapter.

A second class of sentences that Nichols discusses are those that give rise to the *implausible specificity problem*, which Nichols attributes to Hájek in its original form. His own variant goes as follows:

(215) Scenario: I did not strike the match in my hand. The closest worlds where I do are ones where a cluster of neurons in my brain fires differently, sending the right sort of impulse down my arm. There will likely be many different possible neuron-clusters whose firing would have led to the striking of the match. Suppose one of them – call it *xyz* – would have required less departure from actuality than the others, say by involving fewer neurons.

(216) ? If I struck this match right now, it would be a result of neuron cluster *xyz* firing, and not any other.

Nichols agrees with Hájek that (216) sounds implausibly specific – enough so as to render it infelicitous. I am not entirely convinced that I agree with that judgment. Rather, I believe it possible that the oddness of (216) simply arises because it is a backtracking conditional, and backtracking conditionals generally sound somewhat odd, unless they are embedded in a context rich enough, or supported by an extra layer of modality (Arregui, 2005a). Compare the sentence in (217):

(217) If I struck this match right now, it would have to be a result of neuron cluster *xyz* firing, and not any other.

I believe that both (216) and (217) are felicitous in a context where the guiding question under discussion is not *What would have been the result of me striking this match?*, but rather something along the lines of *What would have to have been the case to cause my striking of the match?*. As a result, I do not agree with Nichols (2016) in assuming that both problems need to have a unified solution. However, Nichols proceeds to propose just this, again spelled out in terms of relevance:

(218) $A \boxdot\!\!\rightarrow C$ is true at a context $c$ iff $\forall s \in S_c(f(s) \subset C)$.

In (218), $S_c$ is a set of scenarios $s_i$ relevant in $c$, and $f$ an assignment function that maps each of the scenarios $s_i$ to a set of worlds $W_i$ that witness $s_i$. As Nichols concedes, at least two notions here need further spelling out: that of a scenario, and that of the assignment function $f$ mapping such scenarios to sets of worlds. In order to avoid unnecessary ontological commitments, Nichols simply identifies scenarios with sets of worlds (rather than with a more ontologically loaded term like, say, situations). He then proposes, although tentatively, that the assignment function does not map a scenario to all of its worlds, but rather to a particular subset. Here, he opts for the closest worlds, in the standard sense of closeness being a function of similarity, but also considers other options (e.g. all sufficiently close worlds).

In effect, Nichols' counterfactual first identifies a set of scenarios – roughly, ways of making the counterfactual true – by way of relevance, and then uses similarity to pick, for each such scenario, the worlds most similar to the actual world witnessing that scenario. The resulting set of worlds is his domain of quantification.

### 4.2.3 The notion of relevance

The basic intuitions driving both Lewis's (2016) and Nichols's (2016) account are familiar from our discussion so far. Their data shows that our problems with the standard Lewis-Stalnaker account are well motivated, even beyond the case of counterfactual donkey sentences. What both Lewis

and Nichols propose is that we should go beyond the most similar worlds in some cases, and that the mediating factor in these cases is relevance. I am very sympathetic to this approach, but I am slightly worried about how they employ the notion of relevance: It is very much treated like a black box, making it difficult to assess in which cases a possibility could not be treated as relevant in some way. This is a common problem and may well be solved by giving a more principled account of the notion of relevance, motivated from other areas of the grammar as well.

At the moment, I think it is most worthwhile to highlight the commonalities between this approach and the account of Arregui (2009) which we will discuss shortly: Both Lewis' and Nichols' relevance-based accounts and Arregui's local similarity account give up on requiring maximal similarity for worlds, instead allowing for a larger set of possibilities to be quantified over. In the absence of a principled theory of relevance, however, I find it easier to apply Arregui's account, as it does not rely on this notion. Rather, it achieves a similar effect through a combination of salient facts and law-like regularities. Of course, one can read Arregui's account, and these ingredients, as exactly the more principled account of relevance that we are looking for. I leave this issue for further research.

## 4.3 Causality

### 4.3.1 Causal readings

A number of authors have recently discussed the relationship between causality and counterfactuals (Henderson, 2011; Kaufmann, 2013; Santorio, 2016; Schulz, 2007, 2011). This is, in a way, a reversal of Lewis's (1973) analysis of causation in terms of counterfactuals, instead positing that counterfactuals can be better understood if analyzed in terms of causation. The main insight of this literature, at least for our purposes, is

that counterfactuals are not simply sensitive to similarity understood as "overlap of facts". Rather, we need to impose some additional – causal – structure on the relevant facts and weigh differences in facts differently from deviations from causal laws.

The basic observation common to this strand of literature is that counterfactuals have, as their most salient reading, a causal – or "ontic", in Schulz's (2011) terminology – reading that appears in addition to the standardly assumed epistemic reading. In most cases, these two readings are indistinguishable, as they lead to the same results, but there are a number of cases where they come apart. Schulz (2007) presents the following case as one that clearly demonstrates the ambiguity (here in the variation due to Henderson (2011)):

(219) Suppose an alarm sounds at the docks whenever there is an impending storm. Further suppose, we took the bridge instead of the ferry because we heard the alarm and there was, in fact, a storm.

(220) a. Thank goodness, if the alarm hadn't gone off, we would have taken the ferry and we might have all drowned in the storm.

b. No no no, the alarm always works. If it hadn't gone off, there would have to have been no storm coming.

In this example, (220a) demonstrates the causal reading, while (220b) shows the epistemic reading. That is, according to Schulz (2007), (220a) explores the consequences of a change in the causal structure – an intervention on the sounding of the alarm –, while (220b) explores the conclusions we can draw from learning that the alarm had not sounded.

Henderson (2011) further argues that we are dealing with genuine readings here, as there is a construction that allows for only one of them, blocking the other completely: "if not for"-counterfactuals like (221).

(221) If not for Mary going to the store, we wouldn't have salsa.

(222) If Mary hadn't gone to the store, we wouldn't have salsa.

Henderson points out that (221) and (222) work as paraphrases of one another, but systematically differ in their properties. Specifically, the construction in (221) does not allow for the epistemic reading:

(223) Suppose an alarm sounds at the docks whenever there is an impending storm. Further suppose, we took the bridge instead of the ferry because we heard the alarm and there was, in fact, a storm.

(224) a. Thank goodness, if not for the alarm going off, we would have taken the ferry and we might have all drowned in the storm.

b. # No no no, the alarm always works. If not for the alarm going off, there would have to have been no storm coming.

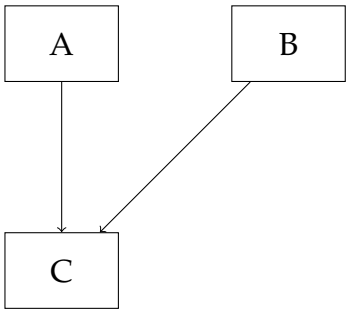
Having established the existence of the causal reading, the challenge for a counterfactual semantics is to account for it. In order to capture the causal reading, Schulz (2007, 2011) proposes to make use of a notion of causal entailment. In the following section, I will sketch Schulz's account.

### 4.3.2 Causal entailment

Schulz (2011) makes use of a notion of situation that slightly deviates from the one that we introduced in Ch. 3.1. Instead of treating situations as concrete objects in the ontology, as Kratzer (1989) does, Schulz uses a slightly more abstract notion in the tradition of Veltman (2005). Using a three-valued logic that adds *u* (for "undefined") to the truth values, assignments of truth values to the complete set of atomic propositions fall into two sets: Those that assign *u* to some proposition are situations, while those that do not assign *u* to any proposition are possible worlds. In this sense, a situation is an incomplete description of a possible world. Some ways of talking about situations that are possible in the Kratzer semantics are not possible under this definition – e.g., an individual cannot directly be part of a Schulz situation – so that these two notions of situation are not straightforwardly interchangeable. In the remainder of this section, I will be using Schulz's notion of situation unless indicated otherwise.

The second technical term that Schulz introduces is that of a *dynamics*, used to represent causal dependencies. A dynamics is a partition on the set of propositional letters that divides them into those that are causally independent, the set $B$ (for "background variable"), and those that causally depend on some other variable, the set $I$ (for "inner variable"). Additionally, a dynamics provides a (two-valued) function that determines the truth values of inner variables based on the truth values of the background variables they depend on[4].

(225)




In (225), $A$ and $B$ are background variables, together determining the value of the inner variable $C$. The figure shows only the causal dependencies, but we need to additionally describe the truth function for $C$, based on $A$ and $B$. For example, if $C$ is 1 if and only if $A$ and $B$ are both 1, then the following information also needs to be encoded in the causal dynamics:

(226)

| $A$ | $B$ | $C$ |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 0 | 0 |
| 0 | 1 | 0 |
| 0 | 0 | 0 |

Given these ingredients, Schulz can formulate a notion of causal entailment based on the recursive application of an operator $\tau$ relative to a dynamics $D$ to a situation $s$. $\tau$ takes a situation $s$, which – by definition – has some undefined values. It then attempts to fill in these values based on the

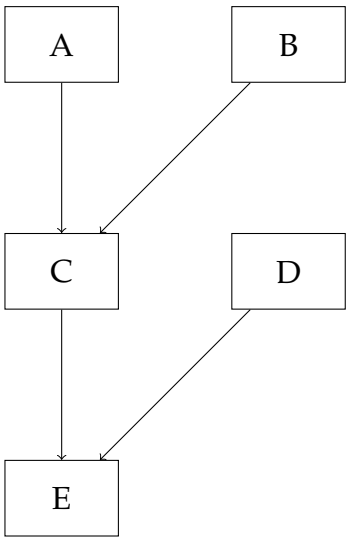
[4] I will limit myself to a somewhat informal introduction of the concepts necessary for our purposes here. For the formal definition of the underlying framework and the necessary constraints on dynamics, e.g. rootedness, see Schulz (2011).

values of other variables, following the causal dependencies specified in $D$. Once application of $\tau$ yields no further change, the resulting situation is returned, and all propositions verified by this resulting situation are taken to be causally implied by the original situation $s$, relative to $D$.

Applying $\tau$ to a situation $s$ and a dynamics $D$ yields the following: The resulting situation $\tau_D(s)$ agrees with $s$ in all the values of the background variables, as well as in all values that are not undefined, or for which there is no truth function defined. For undefined values in $s$ with a defined truth function, $\tau_D(s)$ obtains the value as defined by the truth function.

Take, for example, our dynamics above, specified in (225) and (226). Relative to this dynamics, a situation which sets the values of $A$ and $B$ but leaves $C$ undefined will always causally entail either $C$ or $\neg C$: if $A$ and $B$ are both set to 1, then $\tau$ can compute the value of $C$ as 1, and as 0 otherwise. This generalizes to larger dynamics (and situations) as well. Take the example in (227) and (228) – (229):

(227)




(228)

| $A$ | $B$ | $C$ |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 0 | 0 |
| 0 | 1 | 0 |
| 0 | 0 | 0 |

(229)

| $C$ | $D$ | $E$ |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 0 | 0 |
| 0 | 1 | 0 |
| 0 | 0 | 0 |

Here, the dynamics from (225) is extended: There is another background variable $D$, and another inner variable $E$, which in turn depends on both an inner variable ($C$) and a background variable ($D$) for its value. The truth conditions are simple conjunction again: $E$ is 1 if and only if $C$ and $D$ are both 1, as specified in (229).

A situation where the background variables ($A, B, D$) are all defined and the inner variables ($C, E$) undefined, will causally predict values for both inner variables, given the dynamics above. Here, $\tau$ will have to be applied in two steps though: First, the value of $C$ is computed by the truth table in (228). However, the first application of $\tau$ still yields an undefined value for $E$, because although $E$ is undefined in $s$, there is no defined truth function for it (because $C$ is also undefined in S). Applying $\tau$ to the result of the first application now allows us to compute the value of $E$, since we obtained a defined value for $C$ in the first step and (229) is no longer undefined.

For concreteness, let's take an initial situation $s$ that looks as follows:

(230) A = 1, B = 1, C = u, D = 0, E = u

Applying $\tau$ to (230) yields a situation $s'$ that looks as follows:

(231) A = 1, B = 1, C = 1, D = 0, E = u

How do we obtain (231)? First, $s'$ is required to agree with $s$ in all defined values, that is, in the values of $A$, $B$ and $D$. Then, we check the undefined values $C$ and $E$: For $C$, we have a defined truth table in (228), which for $A = 1, B = 1$ assigns $C$ the value 1 in $s'$. However, for $E$, the truth table in (229) is not defined, because $C$ in $s$ is still $u$. Since we have undefined values left, we apply $\tau$ again and yield the situation $s''$:

(232) A = 1, B = 1, C = 1, D = 0, E = 0

Again, $s''$ agrees with $s'$, as described in (231), in all defined values, i.e. in the values of $A$, $B$, $C$ and $D$. Then we check the undefined value $E$: We now have a defined truth table in (229), because the values of both $C$ and $D$ are defined in $s'$. Together, they set $E$ to 0.

Applying $\tau$ again to $s''$ does not yield any changes, so that $s''$ is the final result – anything in $s''$ is causally entailed by $s$ relative to $D$.

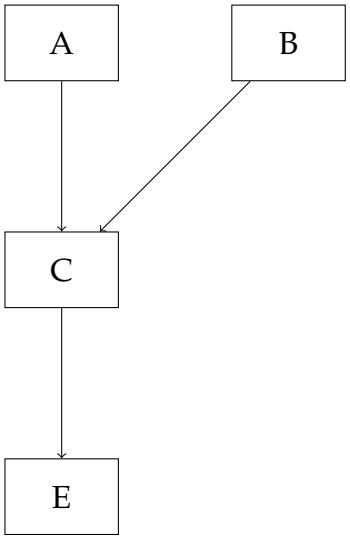
### 4.3.3 Causal counterfactuals

Given this notion of causal entailment, Schulz (2011) gives the following semantics to counterfactual conditionals:

(233) $[\![A > C]\!]^{D,w_0} = 1$ iff $A + P_{w_0} \models_D C$

In addition to the notion of a dynamics $D$ and causal entailment $\models_D$ (which is defined via the operation $\tau$, as described above), this definition relies on a set of premises $P_{w_0}$ and an operation for combining this set with the antecedent $A$. Schulz in turn gives the following definitions for those: $P_{w_0}$ is a *basis* in the sense of Veltman (2005). In causal terms, the basis can be defined rather easily, as the set of all background variables, plus those inner variables whose value violates some causal law as specified in the dynamics. Take, for example, the dynamics in (234) – (236) (which is just the dynamics in (227) with $D$ taken out for simplicity, reducing the relation between $C$ and $E$ to the identity function), and some of the corresponding worlds in (235):

(234)




(235)

| $A$ | $B$ | $C$ |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 0 | 0 |
| 0 | 1 | 0 |
| 0 | 0 | 0 |

(236)

| $C$ | $E$ |
|---|---|
| 1 | 1 |
| 0 | 0 |

(237)

| | $A$ | $B$ | $C$ | $D$ |
|---|---|---|---|---|
| $w_0$ | 1 | 1 | 1 | 1 |
| $w_1$ | 1 | 0 | 0 | 0 |
| $w_2$ | 0 | 1 | 0 | 0 |
| $w_3$ | 0 | 0 | 0 | 0 |
| $w_4$ | 1 | 1 | 0 | 0 |
| $w_5$ | 1 | 1 | 0 | 1 |

The base for $w_0$ simply consists of the background variables $A$ and $B$ – and their truth values, i.e. $\{A, B\}$: Together, they determine the value of $C$, and $C$ determines the value of $D$. The bases for $w_1$, $w_2$ and $w_3$ are equally simple: $\{A, \neg B\}$ and $\{\neg A, B\}$ respectively. That is, as long as the inner variables are all determined by the background variables, the

base only consists of the background variables. Things get slightly more complicated as soon as one of the inner variables does not follow the causal laws. In $w_4$, $A$ and $B$ are both 1, but $C$ is 0 nonetheless (essentially, a Lewisian miracle has occured.) In order to describe the world $w_4$, we need the base $\{A, B, \neg C\}$. The value of $E$, again, follows the causal laws. In $w_5$, both $C$ and $E$ break the causal laws, and so the resulting base is $\{A, B, \neg C, D\}$ – essentially a full description of the world[5].

With this notion of a base $P_{w_0}$ in place, Schulz then defines the revision $P_{w_0} + A$ as the set of maximal subsets of the base that are consistent with the antecedent and the laws in $D$, plus the antecedent itself.

### 4.3.4 Disjunctive antecedents

One particular feature of Schulz's (2011) account is that it is only well-defined for antecedents that are either atomic or can be rewritten as conjunctions of atomic sentences. For disjunctive antecedents, Schulz proposes an interpretation that amounts to high readings, by considering every disjunct separately and universally quantifying over the resulting counterfactuals:

(238) For every formula $\phi$ in the language, $Lit(\phi)$ is the set of sets of literals with the following property: $\{\phi_1, ...\phi_k\} \in Lit(\phi)$ iff $\phi_1 \wedge ... \wedge \phi_k$ is one of the disjuncts in the disjunctive normal form of $\phi$.

(239) $[\![A > C]\!]^{D,w_0} = 1$ iff $\forall S \in Lit(A) \forall B \in S + P_{w_0} : B \vDash_D C$

This, of course, means that in order to generate *low* readings in a causal account, we will need to resort to some extra machinery. One could, e.g., introduce some ranking of the resulting sets $P_{w_0} + S$ and optionally only pick the best one, thus generating a low reading. We would then, again, be using the same mechanisms as proposed in Ch. 2.

[5] Note that in her examples, Schulz (2011) uses very small toy worlds, whereas her definitions seem to assume that the base is indeed defined for an entire possible world. While Schulz's mechanics does scale up to larger worlds, I think there is something intuitively plausible about using only partial causal dynamics, with those variables included that are salient in the context. We will come back to this in 4.5, when we combine Schulz's approach with Arregui's (2009) notion of local similarity.

### 4.3.5 Evaluation

Schulz's (2011) framework adds some important tools to our understanding of counterfactuals. With respect to the problem of this thesis, it does not immediately contribute an advantage: Rather, the problem is reversed, now standardly only generating high readings, with low readings more problematic. However, the additional structure that causal dynamics can add to situations will prove useful in distinguishing between contexts that trigger high and low readings respectively, as I will argue in Ch. 4.5.

## 4.4 Locality

The discussion of counterfactuals' context-dependence has been a long-standing issue. Lewis (1973b) addresses it in his discussion of the famous *Caesar* example (attributed to Quine by Lewis, who in turn attributes it to Goodman in Quine (1960)):

(240) If Caesar had been in command in Korea, he would have used the atom bomb.

(241) If Caesar had been in command in Korea, he would have used catapults.

Intuitively, both (240) and (241) can be judged as true, although they come to very different conclusions. Lewis (1973b) admits that similarity has to be resolved differently in both cases, with different features of the world being more or less important (e.g., Caesar's character as a ruthless military commander in (240) versus his technological knowledge in (241)). But his theory does not spell out the mechanics of this, characteristically keeping similarity a primitive measure that is not analyzed further, its inherent vagueness being relegated to context.

A similar point can be made with the following examples (equally attributed to Goodman by Lewis (1973b)):

(242) If New York had been in Georgia, New York would have been in the south.

(243) If New York had been in Georgia, Georgia would have been in the north.

Arregui (2009) points out that there is a very intuitive way of spelling out the dependency of counterfactuals on particular facts in the world: Instead of making similarity a comparison between entire worlds, as in Lewis (1973b) (a "global" approach in Arregui's terminology), we can also consider a more "local" version of similarity that is based on comparing just those parts of worlds that are relevant to the truth of the counterfactual. Such parts can, e.g., be thought of as situations in the situation-semantic framework already familiar from Ch. 3 (Kratzer, 1989, 2012).

Arregui's account has a second aim that is somewhat orthogonal to the issues discussed in this thesis: She attempts to relate this context-dependence of counterfactuals to the presence of "fake" past tense in counterfactual morphology (Iatridou, 2000). Her account has been located in the family of past-as-past approaches (Dudman, 1983; Grønn and Von Stechow, 2009), as opposed to the past-as-modal line (Iatridou, 2000; Schlenker, 2005; Schulz, 2014). However, it differs crucially from approaches in this family in that it does not treat the evaluation of the counterfactual as the evaluation of an indicative conditional from the perspective of an earlier moment in time. Rather, the function of the past tense in Arregui's account is to pick out certain salient facts (through a past situation) that matter for the evaluation of the counterfactual. I do not intend to take a stand on the debate between past-as-past and past-as-modal accounts here, and there are several issues with Arregui's specific analysis of past tense (Romero, 2014). However, these issues do not directly bear on Arregui's main point regarding the nature of similarity: Similarity is referential, that is, counterfactual conditionals are always true with respect to a particular situation (as opposed to being true of the world simpliciter). Here, I will spell out this insight directly in Arregui's (2009) framework, mostly for expository purposes, but other frameworks could be adapted accordingly, provided we can obtain a salient past situation either from the morphology (as in Arregui's approach) or contextually.

The crucial feature of Arregui's approach for our purposes is its lexical entry for *would*, given in (244), that specifies the behaviour of the counterfactual and spells out how local similarity is determined:

(244) $[\![would]\!]^{w,g}(\mathrm{p})(\mathrm{q})(\mathrm{s}) = 1$ iff
$\{s'_L : s \sqsubseteq_m s'_L \wedge p(s'_L) = 1\} \subseteq \{s'_L : \exists s''_L . s'_L \sqsubseteq s''_L \wedge q(s''_L) = 1\}$

It is easy to see that (244) makes no direct reference to the Lewisian similarity relation[6]. How, then, do we constrain the domain of quantification for the counterfactual in this account? At the heart of Arregui's account lies the idea of the counterfactual being about a particular past situation $s$, the *res* in her terminology, that is provided as a third argument in addition to antecedent and consequent. The counterfactual then quantifies only over those situations $s'_L$ that contain a modal part of the *res*. Modal parthood, the relation $\sqsubseteq_m$, is defined as follows:

(245) $s_i \sqsubseteq_m s_j$ iff $s_i$ has a counterpart in $s_j$

In order to unpack this definition, we will have to take a closer look at counterpart theory.

### 4.4.1 Counterpart theory

Counterpart theory, as used in Arregui (2009), goes back to Lewis (1968, 1971, 1973b, 1983). It is necessitated by the assumption – explicitly made both by Lewis himself, and in Kratzer's (1989) framework which Arregui (2009) inherits – that individuals are world-bound, i.e. they are at most part of one world. Given this assumption, we require a way of identifying distinct individuals in different worlds with one another, that is, a way of finding a given individual's *counterpart(s)* in another world[7].

[6] However, we will have to return to this issue when discussing the modal part-of relation $\sqsubseteq_m$in section 4.4.1.

[7] Note that Elbourne (2013) rejects this limitation, instead treating individuals as existing across worlds. I am not taking a stance on the metaphysical advantages or disadvantages of the respective approaches here. Note, however, as discussed in Ch. 3, that Elbourne still upholds world-boundedness for situations. As we will treat individuals as special cases of situations, I see no particular need to treat them differently in terms of world-boundedness.

As Lewis proposes, the work of identifying counterparts can be done by a counterpart relation based on similarity. Intuitively, individuals count as counterparts of one another if they are similar enough, that is, if they share enough properties to identify them with one another. Admittedly, this is rather vague – as similarity tends to be –, but it does have an advantage over similarity as we have used it so far: While we have very little to offer in terms of intuitions about the similarity of entire worlds, it seems an easier task to identify two individuals from different worlds with one another. There are simply fewer dimensions to keep track of. As such, counterpart theory should not be particularly troubling once we have already accepted the standard Lewis-Stalnaker framework being grounded in similarity. A simple example provided by Arregui (2005b) is the following[8]:

(246) The winner of the last presidential election in the US, Donald Trump, could have been vegan.

According to Arregui, we proceed as follows: We look at individuals in other possible worlds and check whether they are similar enough to count as counterparts of our actual US president, who is an orange-faced individual that eats well-done steak with ketchup[9]. Then, if we find such a counterpart that is vegan, we will consider (246) true, but false otherwise. In a context where similarity with respect to political alliances is important, and where in actuality the president is a strong supporter of the beef industry, we will probably not find a vegan president who is still similar enough to the actual president to count as a counterpart. All counterparts of the actual president will be carnivorous, although they might vary in other dimensions, such as the particular colour of their face.

[8] Arregui's original example is, unfortunately, ambiguous between a *de re* and a *de dicto* reading, which may be misleading. For this reason, I enforce the *de dicto* reading in my variant by naming the individual.

[9] This, again, slightly deviates from Arregui's example, to go with the times.

Arregui (2005b, 2009), following Kratzer (1989) in treating individuals as special cases of situations, extends this notion of counterparthood to situations. Situations are world-bound, but they have counterparts in other worlds. These counterparts are situations that share enough properties to be counted as similar enough, just as in the case of individuals.

### 4.4.2 Counterparts of the *res*

With these assumptions in place, we can return to the discussion of Arregui's semantics for the counterfactual in (244) and (196), repeated below as (247) and (248).

(247) $[\![would]\!]^{w,g}(\mathrm{p})(\mathrm{q})(\mathrm{s}) = 1$ iff
$\{s'_L : s \sqsubseteq_m s'_L \wedge p(s'_L) = 1\} \subseteq \{s'_L : \exists s''_L . s'_L \sqsubseteq s''_L \wedge q(s''_L) = 1\}$

(248) $s_i \sqsubseteq_m s_j$ iff $s_i$ has a counterpart in $s_j$

What we require for the counterfactual's truth is the following: Those situations $s'_L$ that contain a counterpart of the *res* and verify the antecedent $p$ should be a subset of the situations $s'_L$ which are part of a situation $s''_L$ that verifies the consequent and satisfy the set of laws contextually specified in $L$. Note that $L$ is somewhat tucked away in the subscript in Arregui's formula – this should not distract us from its importance.

In identifying counterparts of the *res* we proceed as above: We look for situations that are similar enough to count as such counterparts. However, note that we do not quantify over these counterparts, but rather over lawful antecedent-verifying situations containing them.

There are, then, three degrees of freedom in Arregui's approach: First, counterfactuals can be predicated of different *res* situations, allowing variation in their truth value by identifying the speaker's intended past situation. However, even with the same *res*, counterfactuals can come out differently, depending on the particular counterpart relation employed. For instance, in the example in (246), we assumed a counterpart relation that places more emphasis on similarity with respect to political alliances, but less emphasis on face colour. It is this second variability that we tried to

exploit in Ch. 3.3.2. However, as we saw, without a principled account of the limitations we have to place on counterpart relations, this comes dangerously close to simply stipulating ad-hoc relations that solve the problem at hand. I think that future research may be able to shed more light on the contextual factors that go into choosing a particular counterpart relation, but for the time being we should not expect this factor to be very enlightening. In what follows, it will primarily be the variability due to the third degree of freedom that I will employ to explain the variability in readings for counterfactual donkey sentences: Even when *res* situations and counterpart relations are kept fixed, the set of laws $L$ might be different, leading to different situations counting as lawful extensions.

## 4.5 Low donkeys follow the laws

Let's take stock. If we want to model high readings without committing ourselves to a dynamic semantics for the indefinite, both the accounts of Arregui (2009) and Schulz (2011) will give us what we need[10]: Schulz by treating the indefinite as a disjunction, given the assumption that we universally quantify over disjunctive antecedents in the causal analysis of counterfactuals; Arregui by broadening quantification from the closest worlds to all worlds that are "similar enough" by virtue of containing a counterpart of the *res*. But now the burden is reversed: Instead of having to invoke special machinery to explain the high reading, as we did in Chs. 2 and 3, we are now left without an explanation for the low reading. Have we gained anything? I think yes.

One of the empirical observations of this thesis is that the distribution of high and low readings matches such a reversal: The high reading arises easily whenever we encounter a counterfactual donkey sentence, especially when there is no additional context. It seems to be the standard reading that is available even when a richer context is present, and as such, we would want it to be predicted by our account without any additional

[10] I am now setting aside relevance-based accounts due to the reasons discussed in 4.2.3.

effort. The low reading, on the other hand – although more closely related to what Lewis (1973b) would have predicted in the absence of a closer look at indefinites –, turns out to be the odd case that requires either an identificational construction or a certain amount of context to work.

If we want our theory to explain, rather than just generate the two readings, we will have to find a way to derive the low reading from the role contextual factors play in either of the two theories. What I am going to propose in this chapter is an account that mostly builds on Arregui's (2009) analysis for its general semantics of counterfactuals, making it easy to generate high readings. The reason I choose Arregui's framework as a starting point is that it is entirely independent of the presence of an indefinite in the antecedent of the counterfactual, whereas Schulz's framework is – in a way – more akin to a hard-coded dynamics-style high reading that operates by treating the indefinite as inducing a disjunction about all of its potential referents[11]. I will then proceed to explain how context can serve to exclude certain situations from the domain of quantification, in effect yielding a low reading. In a nutshell, I am going to argue that similarity is dependent on *res* situations, as proposed by Arregui, but that counterfactuals are not neutral with respect to how the antecedent is then introduced into these situations. There are at least two strategies available: First, simply forcing the antecedent via a Lewisian miracle, or, second, making it come about by manipulating an upstream variable in a Schulz-style causal model. Cooperative agents then attempt to make the counterfactual true, which can cause them to choose a low resolution over a high one, with the false high resolution acquiring a non-cooperative "nitpicky" flavor.

Consider the following test case.

(249) John always reads *Dynamics of Meaning* on long train rides. Last week, for some reason he did not read on the train.

(250) If John had read a book on the train ride last week, he would have learned something about donkeys.

[11] For arguments against affording the indefinite such a prominent role in this puzzle see both the discussion in Ch. 2 and Arregui (2005b).

The true low reading of (250) works on the assumption that the most likely book for John to have read is *Dynamics of Meaning*, which teaches its reader something about donkey sentences. There is, however, a false high reading in which *a book* ranges over all possible books, allowing for a reply like (251):

(251) No, that's false. He might have read *Vowels and Consonants* by Ladefoged, and then he wouldn't have learned a single thing about donkeys.

Note that in the context of (249), this response has a somewhat "nit-picky" flavour – it could easily be prefixed with an expression like *technically* or *in principle*. Both readings need to be accounted for by our theory.

Let's assume for simplicity, that there are only three worlds: $w_0$, which is captured in the scenario in (249), $w_1$ in which John reads *Dynamics of Meaning* on the train ride, and $w_2$ in which John reads *Vowels and Consonants* instead. With this setup in mind, let's explore what Arregui's account does and does not predict so far.

First, (250) will be predicated of a particular salient past situation. I will make the simplifying assumption that the most salient situation in the pair of (249) and (250) is one which contains John on his train ride last week, with his disposition to read *Dynamics of Meaning*, but not reading the actual book[12]. Arregui's account then prompts us to find counterparts of this situation in other worlds which are similar enough, but still vary enough for some of them to allow that the antecedent becomes true in them. This means that these counterparts will still contain John and his disposition. Then, we check whether these counterparts can be extended with the antecedent – if they cannot, for example because they are not part of a situation in which John reads a book, they are thrown out. We only quantify over situations that contain both: a counterpart of the *res* and the

[12] Note that the issue of negation is one that quickly gets very complicated in a situation semantics, as Kratzer's (1989) discussion of persistence demonstrates. Here, I only intend to convey that the targeted situation is one that does not contain a book which John is reading (so to speak, an absence of reading), not that it positively contains some negation of John reading.

antecedent. That is, presumably, in our case: situations in which John is on a train ride, has a disposition to read *Dynamics of Meaning* and reads a book. Of these situations, we say that they should all (universally) be extendable into situations in which the consequent holds, that is, in which John learns something about donkeys.

This account predicts that (250) will universally be judged as false, by virtue of $w_2$: It contains such a situation – John is on a train, with his dispositions, reading a book –, but since he is reading *Vowels and Consonants*, he learns nothing about donkeys. That is, the richer context – the additional information about John's dispositions – makes no difference to the counterfactual. In fact, it is judged in the same way that it would have been out of the blue: Minimally, due to the conditional's counterfactuality, we predicate it of a *res* in which John is on a train not reading a book; counterparts will then be John on a train, reading or not reading a book; and finally, extensions with the antecedent will again yield the same situations of John reading either of the two books, just minus his disposition to read *Dynamics of Meaning*.

Our task, then, is to give this additional contextual information a role: In order to yield the low reading under which (250) is true, we need to make sure that the situations in our domain of quantification exclude the situation in $w_2$ in which John reads a book that goes against his reading dispositions. Note that we do *not* need to postulate an ambiguity in the semantics, or a contextually resolved additional mechanism modulating between the two readings: The claim will be that this additional information will *always* restrict the domain of quantification in this way, and that the ambiguity between the two readings simply arises from whether certain regularities are included in the laws in $L$ or not.

### 4.5.1 How to extend a situation

Since the culprit is not the *res* and its counterpart itself, but rather the whole situation quantified over – which includes the antecedent –, we will first need to unpack Arregui's definition of the counterfactual.

(252) $[\![would]\!]^{w,g}(\mathrm{p})(\mathrm{q})(\mathrm{s}) = 1$ iff
$\{s'_L : s \sqsubseteq_m s'_L \wedge p(s'_L) = 1\} \subseteq \{s'_L : \exists s''_L . s'_L \sqsubseteq s''_L \wedge q(s''_L) = 1\}$

(253) $s_i \sqsubseteq_m s_j$ iff $s_i$ has a counterpart in $s_j$

Note that several things are conflated into the modal part-of relation defined in (253). We could rewrite the set of situations to quantify over as follows:

(254) $\{s'_L : \exists s^c . s^c \approx s \wedge s^c \sqsubseteq s'_L \wedge p(s'_L) = 1\}$

This makes it clear that there are four separate conditions, already spelled out in prose in the section above: There needs to be a counterpart of the res ($s_c \approx s$), this counterpart needs to be included in a larger situation ($s_c \sqsubseteq s'_L$), that larger situation needs to verify the antecedent ($p(s'_L) = 1$), and it needs to follow the laws specified in $L$. As discussed in Ch. 3, it is not the first condition, the counterpart relation, that we want to modify. Rather, we will need to have a closer look at the other three conditions. We will especially need to consider what ways of making the antecedent true will be admissible ways, and whether the counterpart can be included in just any antecedent-verifying situation, or whether there are further restrictions.

In its current state, Arregui's semantics is rather lenient: It allows for any situation as long as it contains the *res*-counterpart and a fact that verifies the antecedent[13]. We want this set of situations to be restricted further: It should only contain situations that have a counterpart of the *res* and that verify the antecedent in a way that is somehow sensitive to the content of the *res*. That is, we want to constrain the ways in which the counterpart can be extended to an antecedent-verifying situation. Note that Elbourne's (2005) notion of a "minimal extension" is not going to help us here – both extensions, the one in which John reads *Dynamics* and the one in which he reads *Vowels* in our test case are "minimal" in that they add no extraneous facts beyond those that ensure that the antecedent is

[13] There is, of course, the additional requirement that those situations be "lawful", following the laws contextually specified in $L$. However, this requirement is not spelled out in detail. We will return to this immediately.

verified. Rather, we want to say that there is a way in which John reading *Dynamics* is a "proper" extension of the counterpart, whereas him reading *Vowels* is not. It is that notion of "properness" that we need to define further.

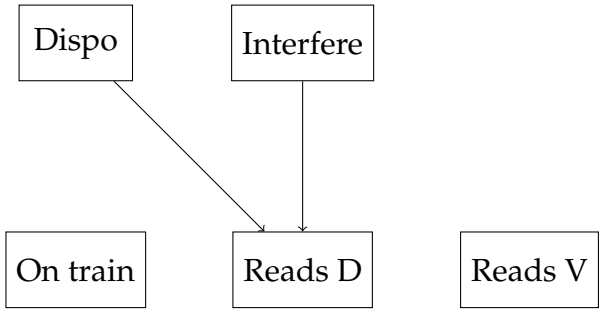
Intuitively, the difference is easy to grasp: The "proper" situation is the one in which John acts according to his dispositions. The "improper" situation is one in which he does not. But is there a way in which we can make this notion more precise? I believe that there is a part of Arregui's semantics that lends itself to this task: the requirement that situations be "lawful", in the sense of following the laws contextually specified in $L$. However, Arregui does not spell out this requirement any further.

### 4.5.2 The role of laws

This is where I think that Arregui's local similarity can productively be extended with some of Schulz's tools. While I will not necessarily subscribe to a particular analysis of counterfactuals in terms of a philosophically controversial term like "causality", I believe that Schulz's notion of a *causal dynamics* is general enough to model the lawlike relations at play in a situation, independent of their particular status. So let's give some more structure to our subscript $L$: In the following, I am going to model at least a subset of the relevant laws as a Schulz-style dynamics. I will again simplify, but in a manner already hinted at in my discussion of Schulz (2011) – I assume that a causal dynamics does not necessarily model the entire world and the laws that hold in it, but can be taken as a representation of a particulary salient set of lawlike relationships.

I will assume that the way to represent the regularities discussed in (249) as a causal dynamics is as follows.

(255)




(256)

| *Dispo* | *Interfere* | *ReadsD* |
|---|---|---|
| 1 | 1 | 0 |
| 1 | 0 | 1 |
| 0 | 1 | 0 |
| 0 | 0 | 0 |

Note specifically that there is a slight shift here: As causal dynamics are deterministic, they do not deal in probabilities. Rather, they can only express dependencies as absolute, and so we will have to simplify John's dispositions to something slightly less complex, namely the truth-functional dependency in (256). Essentially, what (255) and (256) express together is that, in the absence of an interfering factor – such as John forgetting his book at home –, his dispositions cause him to read *Dynamics of Meaning*.

You might, at this point, protest: Surely, this is not the right way to treat dispositions[14]. And I agree – this is no philosophically sound analysis of what John's dispositions are. It is, however, a good model of how we often treat them for conversational purposes. A regularity can easily be awarded the status of a deterministic law and be treated as such as long as nobody raises the possibility of exceptions. And so the difference between low and high readings arises quite naturally: A low reading is one that treats the regularity described in (249) as a law (for conversational purposes), whereas a high reading refuses to accord it such status. Compare how both play out, with (255)–(256) either part of $L$ or not.

[14] Especially so if you are a philosopher who is aware of the host of literature surrounding a concept so controversial (Choi and Fara, 2016). Again, I will attempt to steer clear of unnecessary philosophical commitments.

For the high reading, we can simply take Arregui's account and its predictions as written up in 4.5 and obtain a false reading: As both potential books are indistinguishable from each other for the purposes of extending the counterpart of the *res* – they both require us to assume a small miracle that makes John read a book –, we need to take both extensions into account, rendering the counterfactual false. However, for the low reading we will now assume that the situations to be quantified over are constrained by $L$, specifically, by an $L$ that incorporates the causal dynamics in (255)–(256). We will interpret Arregui's "being a lawful extension" in a Schulzian way, that is, as "can be derived from the non-extended situation by causal entailment." (where, again, no claims about causality are intended, and "causal" is simply to be read as "law-like").

Starting from a *res* that has counterparts in which John is on a train, together with his dispositions, but no book yet, we can now simply follow the causal model. Instead of requiring a Lewisian miracle, we simply assume that – other than in the actual world – nothing interferes with John's dispositions, such as him forgetting his book. But then, given our causal model in (255)–(256), it is causally entailed by John's dispositions and the lack of interference that he reads *Dynamics of Meaning*.

More formally, consider the following setup:

(257) $\forall s^c \approx res$: Dispo$(s^c) = 1$, Interfere$(s^c) = u$, Reads-D$(s^c) = u$, Reads-V$(s^c) = u$

I take (257) to be an adequate description of the counterparts of the *res*: We know that John still has his dispositions, but there is nothing specifying either any interference or reading. For the high reading, where the causal dynamics in (255)–(256) is not part of $L$, changing the value of Interfere will not make any difference. In order to satisfy the antecedent, we will have to directly set either Reads-D or Reads-V to 1. This requires a small miracle, but it does so in both cases, making them indistinguishable. Under these circumstances, we have to evaluate the counterfactual with respect to both options.

For the low reading, there is a third option: setting INTERFERE to 0. This, together with (255)–(256) will causally entail that READS-D is also set to 1. Under these circumstances, we can distinguish READS-D from READS-V, and as cooperative hearers, we will use this option to yield a true (low) reading for the counterfactual.

This way, we can explain how the low reading arises in contexts that are rich enough to support it: Such contexts provide generalizations that can be treated as deterministic laws for the purposes of conversation – we read the books we are disposed to read, we inherit the donkeys that have been written into a will. We also obtain an explanation for the "nitpicky" flavour of high readings in such contexts – they require us to mistrust our generalizations, and as such are similar to skeptical arguments against inductive generalizations. In the same way we would have to agree that someone who claims that it is not certain the sun will rise tomorrow is technically right, we will have to agree with someone who brings up a high reading. But for most purposes, we will still act as if our generalizations are solid, low readings obtain and the sun rises every morning[15].

Note that this also explains the low reading in identificational sentences like this one from Ch. 1:

(258) If Alex had been married to a girl from his class, it would have been Sue.

The question (258) is addressing in the discourse is not one about the results of Alex marrying – rather, it explicitly asks about the regularities governing his marrying behaviour. This forces us to entertain such regularities in answering the question, and therefore yields a low reading.

[15] Note though that the sceptic may well have a point if we were too careless in our generalizations; in that way, he differs from Kratzer's (1989) lunatic, whose nitpickiness is much more fundamental.

## 4.6 Summary and discussion

In this chapter, I have proposed to approach the problem of high and low readings from the opposite direction: Using a framework that standardly has the counterfactual quantify over more possibilities than the most similar worlds (Arregui, 2009) easily allows us to derive the high reading. This matches our intuitions that the high reading arises in the absence of further contextual information. Low readings arise when context optionally serves to restrict the possibilities quantified over. I have suggested that this restriction comes about through regularities that are salient in the context, and have proposed modelling these regularities with the tools from causal models (Schulz, 2011).

In doing so, we exploit the fact that Arregui's framework is quite flexible in what readings it can generate, based on the *res* situation picked out by the past tense, the counterpart relation employed in identifying counterparts of the *res*, and the lawlike regularities employed in restricting the set of counterparts. This power, unfortunately, carries with it the danger of overgeneration. Note that under the analysis in Ch. 2, we only predict two possible readings (per indefinite): One in which we consider all the potential referents of the indefinite, and one in which we only consider the closest ones. Here, however, we open up the possibility of intermediate readings: Given that Arregui's framework standardly provides the high reading, restricting this set by regularities does not necessarily yield only the most similar candidates. Rather, depending on the regularities that are considered relevant, we may also be able to yield intermediate readings where we target particular subsets of the high reading. Unfortunately, the existence of such readings is hard to test empirically. Since we have no way of reliably inducing a similarity measure in a context, it is much harder to construct examples that rely on gradient similarities (with different donkeys being more likely to be owned due to different regularities,

with different degrees of relevance), rather than on simple binary contrasts as in the examples above (where a donkey is either relevant or not). If such examples could be constructed, they would then allows us to more reliably distinguish between the proposals in Ch. 2 and Ch. 4.

## 4.7 From high readings to Sobel sequences

With these results in mind, let me briefly turn towards a somewhat more speculative part. Here are some observations on how our data relates to other phenomena, suggesting that there might be underlying mechanisms that are common to both parts of the grammar, or even the possibility of a unified analysis. Specifically, I am going to compare high and low readings to conditionals with a backtracking resolution, and to Sobel sequences.

Making the antecedent true by changing a "causally" upstream variable is reminiscent of conditionals with a backtracking resolution (Arregui, 2005b). Note that these are distinct from backtracking conditionals in the sense of Lewis (1979), which are often – though not necessarily – marked with an additional layer of modality (Arregui, 2005a; Ward, 2014). It can be shown that low counterfactual donkey sentences are not proper backtrackers, as they fail Arregui and Biezma's (2016) test: Backtracking counterfactuals "cannot be understood as answering a question regarding how the consequent would have been brought about", as shown in Arregui and Biezma's (2016) example in (259), but low readings clearly can, as shown in (260):

(259) a. $\text{QUD}_{implicit}$: In what circumstances would the plane have to have departed at 1:00?

b. # If the plane had arrived at 2:00, it would have to have departed at 1:00.

(260) a. $\text{QUD}_{implicit}$: In what circumstances would John have learned something about donkeys?

b. If John had read a book on the train, he would have learned something about donkeys.

Instead, low readings seem to be more alike to what Arregui (2005a) calls "counterfactuals with a backtracking resolution". These are counterfactuals like the following (262) in the context of Lewis's (1979) famous original backtracking example in (261):

(261) Jim and Jack quarreled yesterday, and Jack is still hopping mad. We conclude that if Jim asked Jack for help today, Jack would not help him. But wait: Jim is a prideful fellow. He never would ask for help after such a quarrel; if Jim were to ask Jack for help today, there would have to have been no quarrel yesterday. In that case Jack would be his usual generous self. So if Jim asked Jack for help today, Jack would help him after all.

(262) If Jim asked Jack for help today, Jack would help him.

According to Arregui, (262) gets interpreted as a strengthened version of the conditional along the lines of (263), and indeed a similar paraphrase seems available for our low readings, as shown in (264):

(263) If there had been no quarrel and Jim asked Jack for help, Jack would help him.

(264) If John had not forgotten his book and John had read a book on the train, he would have learned something about donkeys.

Arregui does not analyze conditionals with a backtracking resolution further, but notes that they seem to be in some way related to Sobel sequences like the ones discussed in Ch. 2.2. Indeed it is striking that there is a great deal of similarity between the problems discussed here (and in 4.2) and Sobel sequences. Consider the following pair of sentences:

(265) a. If John had gone to the donkey market yesterday, he would have seen Platero.

b. That's false. If John had gone to the donkey market yesterday, he might have gone by night, and then he wouldn't have seen Platero.

There are two potential perspectives on this: Coming from the discussion in this thesis, I can think of (265a) as an intended low reading (the most likely way for John to go to the market is by day), with the speaker in (265b) going for a high reading over potential market visiting times (without an overt indefinite). Note that the high reading becomes more easily accessible if we introduce an indefinite:

(266) If John had gone to the donkey market at some point during the past 24 hours, he would have seen Platero.

(266) is more easily interpreted as entailing that any time point during the past 24 hours is a time point at which John would have seen Platero than (265b). Similarly, adding a richer context can push us towards the low reading:

(267) John usually goes to the donkey market in the afternoon. So if John had gone to the donkey market yesterday, he would have seen Platero.

Here, we are prompted to treat John's disposition as law-like and consequently only quantify over afternoon visits to the market. But now consider this second perspective: We can also simply treat (265a) and (265b) as elements in a Sobel sequence, as in (268) below.

(268) If John had gone to the donkey market yesterday, he would have seen Platero. But if he had gone to the donkey market yesterday during the night, he wouldn't have seen Platero.

Functionally, a high reading and the second element of a Sobel sequence are also alike: The high reading arises when multiple possibilities of satisfying the antecedent are salient, either contextually or through lexical material like an indefinite or a disjunction. The Sobel continuation essentially forces us to consider such possibilities, even if we have previously ignored them. And in the same way we can object to a high reading by affording law-like status to a low reading, we can object to a Sobel continuation:

(269) a. If John had gone to the donkey market yesterday, he might have gone by night, and then he wouldn't have seen Platero.

b. Sure, in principle, but John usually goes to the donkey market in the afternoon.

(270) a. If John had gone to the donkey market yesterday, he would have seen Platero. But if he had gone to the donkey market yesterday during the night, he wouldn't have seen Platero.

b. Sure, in principle, but John usually goes to the donkey market in the afternoon.

Conceptually – albeit not yet formally – I think it makes sense to equate some of the notions from our dynamic proposal in Ch. 2 and our current proposal in Ch. 4: The sufficiently similar worlds that Arregui's counterfactual quantifies over – characterized by containing a counterpart of the *res* – can be thought of as the outer rim of the modal horizon. A high reading targets all possibilities that have been made salient, either explicitly or contextually. A low reading targets only those possibilities that match contextually provided law-like relations. But note that the target of a high reading is also context-dependent: In a way, we can relate shifting standards of counterparthood with a shifting modal horizon. And a very wide modal horizon can always be reigned in by pointing out regularities that exclude some of the possibilities under consideration.

# Chapter 5

# Conclusions

At the beginning of this dissertation, we set out to explore the puzzle of high and low readings in counterfactual donkey sentences, hoping to learn something about the way that similarity works in the standard similarity-based framework (Lewis, 1973b). Have we succeeded? And where do we go next? Let me briefly attempt to give an overview.

## 5.1 Dynamic approaches

The dynamic approaches to counterfactual donkey sentences are a conservative extension of the standard semantics. Low readings are generated in the usual way, while high readings can be yielded by partializing the Lewisian similarity relation on the basis of the assignments that are standardly generated by the indefinite in a dynamic semantics for donkey sentences. Counterfactual donkey sentences, from this perspective, really are something that falls out from the interaction between counterfactuals and donkey sentences, both construed in relatively standard ways. Of course, there is a lively debate as to what constitutes the standard approach, and as I have argued in Ch. 2, the particular implementation of the Lewisian semantics we should be employing is von Fintel's (1999) dynamic strict variant.

In Ch. 3, I have asked the same question – that is, which theories do we want to conservatively combine with one another? – with respect to donkey sentences. However, here I conclude that D-type theory does not provide all the resources we need to account for our data. This sheds light on the relationship between dynamic semantics and D-type theory in more general terms as well: While dynamic solutions can often be reconstructed in a D-type framework, a crucial feature of dynamic semantics is that it keeps track of the lexical material that has introduced assignments via co-indexation. D-type theory, on the other hand, discards this information, leading to problems both with the familiar "bishop" problem and with counterfactual donkey sentences. Whether there is a solution to this problem that does not turn D-type theory into a notational variant of dynamic semantics is an open question. For the time being, I think the best framework for constructing a conservative approach to counterfactual donkey sentences is one that combines von Fintel's semantics for counterfactuals with a dynamic semantics for donkey sentences. In this dissertation, we have used classical DPL (Groenendijk and Stokhof, 1991), but in future research we might also want to consider more recent developments in dynamic semantics (Brasoveanu, 2008, 2013).

In further pursuing this conservative line, there are some desiderata. First, the dynamic approaches, in explaining high readings as a byproduct of the anaphoric potential of indefinites, clearly tie their analysis to the presence of indefinites in the antecedent of counterfactual donkey sentences. However, as I point out in Ch. 4, there are some plausible candidates that display the same behaviour in the absence of indefinites. Accounting for this data would require us to posit covert indefinites in these sentences. This is not unheard of – for example, AnderBois (2011) makes such a move to account for cases of sluicing that in a similar way seem to depend on the presence of an indefinite. However, whether covert indefinites are actually present remains an empirical question, and as such can only be answered by further research. AnderBois' research may provide a suitable starting point. However, it is clear that any account that uses covert indefinites will

have to tightly constrain their distribution, and it is not entirely clear to me how the constraints proposed in AnderBois (2011) – which all depend on particular aspects of the construction he is trying to account for – could be translated into constraints that fit our phenomenon at hand.

In addition to this independent account of covert indefinites, we would also want a pragmatic theory that hooks up the dynamic mechanics for generating high and low readings – the contextual variable that marks indefinites as either high or low – with the contexts in which we actually observe high and low readings. I provide no such theory here, but I believe that the basic insights from Ch. 4 would be the right point to start formulating it, connecting Schulz's (2011) causal models to the dynamic apparatus, thus moving to a richer representation of the context than we currently employ.

As a last issue, let me also point out that the phenomenon of high and low readings seems to extend beyond counterfactuals, also showing up in indicative conditionals that contain some other form of modality, e.g. deontic indicative donkey sentences. As we discuss in Walker and Romero (2016), the basic mechanics of this account for counterfactual donkey sentences can be generalized to any other ordering source. However, it is most likely that there is some more unexplored empirical territory in this direction.

## 5.2 Local similarity approaches

The local similarity approach I propose in Ch. 4 is a more radical departure from the standard semantics. In that chapter, I argue that high and low readings are not necessarily an artifact of the interaction between indefinites and counterfactuals. Rather, they are inherent to the semantics of counterfactuals. I argue that in moving from the standard similarity semantics (Lewis, 1973b) to Arregui's (2009) local similarity framework, we capture an important empirical point that the dynamic approaches miss: Where context does not provide additional information about the poten-

tial referents of an indefinite, the high reading standardly arises. It is only through additional information brought in by the context that low readings appear. In Ch. 4, I propose that this additional information consists of regularities that can be represented as causal models (Schulz, 2011).

Again, there are some desiderata that come with pursuing this account further. First, we need to make sure that the account does not overgenerate readings. This will require further empirical research into the range of available readings, but it will also require us to work out independent motivations for some of the tools employed here. The way regularities play a role in this account should be mirrored by the role that they play in other constructions. It would also be helpful to clarify their connection to relevance-based accounts as the ones by Lewis (2016) and Nichols (2016) discussed in Ch. 4.

Note also that in decoupling our account for high and low readings from the semantics for donkey sentences, we have not rid ourselves of the need to account for the anaphoric potential of indefinites. Rather, the semantics proposed in Ch. 4 will be handling the counterfactual side, with one of the standard semantics for donkey sentences handling the anaphoric side. For the reasons outlined above, I believe that this standard semantics should again be a dynamic semantics.

Last but not least, as with the dynamic approaches, we would also want to explain the availability of high and low readings in indicative donkey sentences with some other form of modality. The mechanics of dynamic approaches generalize quite easily to those cases, as discussed in Walker and Romero (2016). Since the mechanics in local similarity approaches are conceptually more closely tied to the domain of counterfactuals, we would need to consider how they might be extended.

# Bibliography


AnderBois, Scott (2011). "Issues and alternatives". PhD thesis. University of California, Santa Cruz.

Arregui, Ana (2005a). "Layering modalities". Ms., University of Ottawa.

— (2005b). "On the accessibility of possible worlds: The role of tense and aspect". PhD thesis. University of Massachusetts, Amherst.

— (2008). "Some remarks on domain widening". In: *Proceedings of the 27th West Coast Conference on Formal Linguistics*, pp. 45–53.

— (2009). "On similarity in counterfactuals". In: *Linguistics and Philosophy* 32.3, pp. 245–278.

Arregui, Ana and María Biezma (2016). "Discourse Rationality and the Counterfactuality Implicature in Backtracking Conditionals". In: *Proceedings of Sinn und Bedeutung 20*. Ed. by Nadine Bade, Polina Berezovskaya, and Anthea Schöller, pp. 91–108.

Barker, Chris and Chung-chieh Shan (2008). "Donkey anaphora is in-scope binding". In: *Semantics and Pragmatics* 1, pp. 1–46.

Barwise, Jon and Robin Cooper (1981). "Generalized quantifiers and natural language". In: *Philosophy, Language, and Artificial Intelligence*. Springer, pp. 241–301.

Bennett, Jonathan (2003). *A philosophical guide to conditionals*. Oxford University Press.

Bird, Steven, Ewan Klein, and Edward Loper (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media.

Brasoveanu, Adrian (2008). "Donkey pluralities: plural information states versus non-atomic individuals". In: *Linguistics and Philosophy* 31.2, pp. 129–209.

— (2013). "The grammar of quantification and the fine structure of interpretation contexts". In: *Synthese*, pp. 1–51.

Brasoveanu, Adrian and Jakub Dotlacil (2016). "Donkey anaphora: Farmers and bishops". In: *The Companion to Semantics*. Wiley.

Büring, Daniel (2004). "Crossover situations". In: *Natural Language Semantics* 12.1, pp. 23–62.

Choi, Sungho and Michael Fara (2016). "Dispositions". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Metaphysics Research Lab, Stanford University.

Cooper, Robin (1979). "The interpretation of pronouns". In: *Syntax and Semantics* 10, pp. 61–92.

Déchaine, Rose-Marie and Martina Wiltschko (2002). "Decomposing pronouns". In: *Linguistic Inquiry* 33.3, pp. 409–442.

Dekker, P.J.E. (1993). "Transsentential meditations: Ups and downs in dynamic semantics". PhD thesis. University of Amsterdam.

DeRose, Keith (1999). "Can it be that it would have been even though it might not have been?" In: *Noûs* 33.13, pp. 385–413.

Dudman, VH (1983). "Tense and time in English verb clusters of the primary pattern". In: *Australian Journal of Linguistics* 3.1, pp. 25–44.

Elbourne, Paul (2005). *Situations and Individuals*. MIT Press.

— (2009). "Bishop sentences and donkey cataphora: A response to Barker and Shan". In: *Semantics and Pragmatics* 2.1, pp. 1–7.

— (2010). "On bishop sentences". In: *Natural Language Semantics* 18.1, pp. 65–78.

— (2013). *Definite descriptions*. Oxford University Press.

Evans, Gareth (1977). "Pronouns, quantifiers, and relative clauses (I)". In: *Canadian Journal of Philosophy* 7.3, pp. 467–536.

— (1980). "Pronouns". In: *Linguistic Inquiry* 11.2, pp. 337–362.

Fauconnier, Gilles (1975). "Polarity and the scale principle". In: *Proceedings of the Chicago Linguistic Society*. Vol. 11, pp. 188–199.

Fauconnier, Gilles (1978). "Implication reversal in a natural language". In: *Formal semantics and pragmatics for natural languages*. Springer, pp. 289–301.

Fine, Kit (1975). "Critical notice". In: *Mind* 84.1, pp. 451–458.

Geach, P. T. (1962). *Reference and Generality*. Cornell University Press.

Groenendijk, Jeroen and Martin Stokhof (1991). "Dynamic Predicate Logic". In: *Linguistics and Philosophy* 14.1, pp. 39–100.

Grønn, Atle and Arnim Von Stechow (2009). "Temporal interpretation and organisation of subjunctive conditionals". Ms. U. Oslo.

Heim, Irene (1982). "The semantics of definite and indefinite noun phrases". PhD thesis. University of Massachusetts at Amherst.

— (1984). "A note on polarity sensitivity and downward entailingness". In: *Proceedings of NELS*. Vol. 14, pp. 98–107.

— (1990). "E-type pronouns and donkey anaphora". In: *Linguistics and Philosophy* 13.2, pp. 137–177.

Henderson, Robert (2011). ""If not for" Counterfactuals: Negating Causality in Natural Language". PhD thesis. University of California, Santa Cruz.

Iatridou, Sabine (2000). "The grammatical ingredients of counterfactuality". In: *Linguistic Inquiry* 31.2, pp. 231–270.

Ippolito, Michela (2006). "Remarks on only". In: *Proceedings of SALT*. Vol. 16, pp. 77–87.

— (2008). "On the meaning of only". In: *Journal of Semantics* 25.1, pp. 45–91.

Kadmon, Nirit (1987). "On the unique and non-unique reference and asymmetric quantification. PhD. University of Massachusetts 1987". PhD thesis. University of Massachusetts at Amherst.

Kadmon, Nirit and Fred Landman (1993). "Any". In: *Linguistics and Philosophy* 16.4, pp. 353–422.

Kamp, Hans (1981). "A theory of truth and semantic representation". In: *Formal semantics. The essential readings*, pp. 189–222.

Kaufmann, Stefan (2013). "Causal premise semantics". In: *Cognitive Science* 37.6, pp. 1136–1170.

Khoo, Justin (2016). "Backtracking Counterfactuals Revisited". In: *Mind*.

Kratzer, Angelika (1979). "Conditional necessity and possibility". In: *Semantics from different points of view*. Springer, pp. 117–147.
— (1989). "An investigation of the lumps of thought". In: *Linguistics and Philosophy* 12.5, pp. 607–653.
— (2012). *Modals and Conditionals*. Oxford University Press.
— (2016). "Situations in Natural Language Semantics". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Spring 2016. Metaphysics Research Lab, Stanford University.
Kroll, Nicky (2008). "On bishops and donkeys". In: *Natural Language Semantics* 16.4, pp. 359–372.
Ladusaw, William A (1980). "Polarity Sensitivity as Inherent Scope Relations." PhD thesis. University of Texas.
Lewis, David K (1968). "Counterpart theory and quantified modal logic". In: *The Journal of Philosophy* 65.5, pp. 113–126.
— (1971). "Counterparts of persons and their bodies". In: *The Journal of Philosophy* 68.7, pp. 203–211.
— (1973a). "Causation". In: *The Journal of Philosophy* 70.17, pp. 556–567.
— (1973b). *Counterfactuals*. Blackwell, pp. 278–281.
— (1975). "Adverbs of quantification". In: *Formal semantics. The essential readings*, pp. 178–188.
— (1979). "Counterfactual dependence and time's arrow". In: *Noûs* 13, pp. 455–476.
— (1983). "Individuation by Acquaintance and by Stipulation". In: *The Philosophical Review* 92.1, pp. 3–32.
Lewis, Karen S (2016). "Elusive counterfactuals". In: *Noûs* 50.2, pp. 286–313.
Nichols, Cory (2016). "Rethinking similarity". Ms. Princeton University.
Patel-Grosz, Pritty and Patrick Grosz (2017). "Revisiting pronominal typology". In: *Linguistic Inquiry* 48.2, pp. 259–297.
Placek, Tomasz and Thomas Müller (2007). "Counterfactuals and Historical Possibility". In: *Synthese* 154.2, pp. 173–197.
Postal, Paul (1966). "On so-called pronouns in English". In: *Monograph series on language and linguistics* 19, pp. 177–206.

Quine, Willard Van Orman (1960). *Word and object*. MIT Press.

Romero, Maribel (2014). "'Fake Tense' in Counterfactuals: A Temporal Remoteness Approach". In: *The Art and Craft of Semantics: A Festschrift for Irene Heim*, pp. 47–63.

Root, Rebecca Louise (1986). "The semantics of anaphora in discourse". PhD thesis. University of Texas.

Santorio, Paolo (2016). "Interventions in premise semantics". In press at Philosophers' Imprint.

Schlenker, Philippe (2005). "The lazy Frenchman's approach to the subjunctive". In: *Romance Languages and Linguistic Theory 2003*, pp. 269–309.

Schubert, Lenhart K and Francis Jeffry Pelletier (1987). "Problems in the Representation of the Logical Form of Generics, Plurals, and Mass Nouns". In: *New Directions in Semantics*. Ed. by Ernest Lepore. Academic Press, pp. 385–451.

Schulz, Katrin (2007). "Minimal models in semantics and pragmatics: Free choice, exhaustivity, and conditionals". PhD thesis. University of Amsterdam.

— (2011). "If you'd wiggled A, then B would've changed". In: *Synthese* 179.2, pp. 239–251.

— (2014). "Fake Tense in conditional sentences: a modal approach". In: *Natural Language Semantics* 22.2, pp. 117–144.

Schwarz, Florian (2009). "Two types of definites in natural language". PhD thesis. University of Massachusetts at Amherst.

Schwarz, Wolfgang (2013). "Counterfactuals with unspecific antecedents". Ms. U. Edinburgh.

Slote, Michael A (1978). "Time in counterfactuals". In: *The Philosophical Review* 87.1, pp. 3–27.

Stalnaker, Robert C. (1968). "A Theory of Conditionals". In: *Americal Philosophical Quarterly* Monograph Series, 2, pp. 98–112.

Tichý, Pavel (1976). "A counterexample to the Stalnaker-Lewis analysis of counterfactuals". In: *Philosophical Studies* 29.4, pp. 271–273.

Tooley, Michael (2003). "The Stalnaker-Lewis approach to counterfactuals". In: *The Journal of Philosophy* 100.7, pp. 371–377.

Van Rooij, Robert (2006). "Free Choice Counterfactual Donkeys". In: *Journal of Semantics* 23.4, pp. 383–402.

Veltman, Frank (2005). "Making counterfactual assumptions". In: *Journal of Semantics* 22.2, pp. 159–180.

Von Fintel, Kai (1999). "NPI licensing, Strawson entailment, and context dependency". In: *Journal of semantics* 16.2, pp. 97–148.

— (2001). "Counterfactuals in a dynamic context". In: *Current Studies in Linguistics Series* 36, pp. 123–152.

Walker, Andreas (2014). "A D-type Theory Solution to the Proportion Problem". In: *Proceedings of the ESSLLI 2014 Student Session*, pp. 165–176.

Walker, Andreas and Maribel Romero (2015). "Counterfactual donkey sentences: A strict conditional analysis". In: *Proceedings of Semantics and Linguistic Theory (SALT) 25*, pp. 288–307.

— (2016). "High and low readings in indicative donkeys". In: *Proceedings of Sinn und Bedeutung 20*, pp. 761–778.

Wang, Y. (2009). "Counterfactual Donkey Sentences: A Response to Robert van Rooij". In: *Journal of Semantics* 26.3, pp. 317–328.

Ward, Kaeli Shannon (2014). "Backtracking and have to: Maintaining a Unified Analysis of Conditionals". PhD thesis. UC Los Angeles.