Counterfactuals and the Logic of Imaginative Content

MSc Thesis *(Afstudeerscriptie)*

written by

Dean McHugh
(born 9 November, 1993 in Dublin, Ireland)

under the supervision of **Dr Katrin Schulz**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam*.

Date of the public defense:	**Members of the Thesis Committee:**
17 September, 2018	Dr Maria Aloni
	Dr Peter Hawke
	Dr Katrin Schulz (supervisor)
	Prof Yde Venema (chair)

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Abstract

Much of what we say aims to spark the imagination rather than inform. Imperatives to imagine such as "Imagine you are eating coconut ice cream" as well as counterfactual antecedents such as "If we hadn't met at the gelato bar..." serve to raise hypothetical contexts. This essay is about the logic of such imaginative talk. In particular, we analyse how three words—*and*, *or*, and *not*—evoke imagined scenarios. The resulting semantics, which we axiomatise, combines contemporary developments in fine-grained semantics (principally, truthmaker and inquisitive semantics), although the logic turns out to be something new.

At the heart of the present account is the construction of a semantic object representing the 'ways of imagining' conjunctions, disjunctions and negations true. This object is designed to express the 'imaginative content' of a sentence, which, we show, differs markedly from its informative content.

All of this is motivated by the idea that to understand the meaning of the words *and*, *or* and *not* in counterfactual antecedents we must understand their contribution in evoking hypothetical contexts. Since counterfactuals are modal contexts, inviting us to look beyond the actual world, we therefore present and strongly axiomatise some 'imaginative' modal logics that serving as the basis for a logic of counterfactuals in which counterfactual antecedents are interpreted in terms of their imaginative content. On this approach, a counterfactual is true just in case the consequent holds under every way of imagining the antecedent true.

Our final substantive chapter highlights one model in particular of the counterfactual logic we present. In this model—which it is tempting to call the standard model—a sentence's imaginative content is understood as a list of ways to change the atomic facts of actual world to imagine the sentence true, and a counterfactual is true just in case the consequent holds under every way of changing the atomic facts of the actual world to imagine the sentence true. We show that this semantics of counterfactuals in terms of atomic change validates a number of desirable rules of inference of counterfactual logic while avoiding some undesirable ones.

We conclude with some puzzles for the reader.

Contents

Acknowledgements

First and foremost, to my supervisor, Katrin Schulz. Thank you. Thank you for being always fair and always right; for believing in me and working tirelessly to help me find a place in academia. I look forward to the next four years, which you did so much to help secure, with you as a role model academic. To my mentor, Maria Aloni, for dispensing the best advice and keeping track of practicalities. To Luca Incurvati, who taught me in eight weeks of basic logic what could have taken a lifetime to learn. And to the ILLC community as a whole, for offering the best research environment I have ever come across, and who knows, if Jonathan and Morwenna predict right, might never leave.

Thanks to you, Jonathan and Morwenna, for introducing me to a tantalising slice of semantics during those long yogurt, bagel and uijfertoffel evenings. (*A Mhorwenna, ní bheidh do leithéid ann arís... seachas i gCalifornia, 2021!*) To Grzegorz, for keeping the whole class sane with regular foosball breaks and introducing me to Botterweg 14. To Hana and Mina, for always having insightful, clear judgements to hand; to Saúl, for always seeing the best in people and for leaving me in awe of your math mode; to Rachael, Sam and Zhuoye, for the games this year and the games to come.

And finally, to my parents Vivienne and Michael—the how and the why. This thesis is dedicated to you. Mam, for your wisdom; and Dad, your wonder.

Human rationality depends on imagination. People have the capacity to be rational at least in principle because they can imagine alternatives.... The principles that guide the possibilities people think of are principles that underpin their rationality.

— Ruth Byrne

The Rational Imagination (2007, 29)

Chapter 1

Introduction

The mind is an imaginative organ. So much of what we reason, speak and argue about is beyond actuality. We wonder why the dinosaurs went extinct, and whether humans will someday follow them. We weigh up whether we will regret moving to another country, or whether to start a family. So much of what actually keeps us up at night are the hypotheticals—the everyday *what ifs* whose solutions can only be imagined.

We might in turn call imagination "the organ of meaning" (C. S. Lewis, 1939, 157). We often raise imagined scenarios just by speaking. Lying in bed at night, imagining how things might turn out, it seems we do not just simulate feelings, sounds, images and other sensory content. We also internally *speak* sentences to ourselves. This language that we use to structure our imagination is the same language a doctor speaks when explaining a patient's options, and it is the same language the undecided voter uses, running policy outcomes through their head before walking into the polling station. Just by talking, we can create imagined scenarios in the minds of others and ourselves. There is a language, then, a fragment of natural language, we use to imagine hypothetical scenarios. And as a language, it has a grammar, rules of inference, and assigns meanings to its sentences. There is, in a word, a *logic* of imaginative content.

This essay is about the logic of imaginative content. It is about how we use sentences in the language of hypotheticals to raise and discuss imagined scenarios. Specifically, we are interested in how three words—*and*, *or* and *not*—structure our reasoning about hypothetical contexts. We illustrate in (1) some of the kinds of sentences we consider in this essay.

(1) a. Set one more place for dinner, in case Mina or Rachael come.
b. If Rob and Nora had not both come to the party, they might have gone climbing instead.
c. You would drink the tea had I put milk, or milk and hemlock in it.

We will not take sides on what it cognitively or neurologically means to imagine a sentence true. For every theory has its primitives: the primitives here are the ways of imagining atomic sentences true and the ways of imagining them false. There is much psychological work (e.g. Tversky and Kahneman, 1974; Byrne, 2007) charting the 'joints' in mental representations that guide what alternatives to reality we select when constructing a hypothetical context. This literature is primarily concerned with the imagination of what a logician would call atomic sentences (e.g. "if you had turned left, ..."), without considering how these basic forms combine under the logical connectives.

We therefore take as our starting point this basic correspondence between language and cognition, whereby atomic sentences and their negations are assigned some 'imaginative content'. Our logic of imaginative content will then describe how these basic forms interact with negation, disjunction and conjunction to produce complex imaginative contents, expressed by sentences of composite form—as we saw, for example, in (1).

The heart of the present account is the construction of semantic object describing the 'ways of imagining' a given sentence true, where sentences are built up from atomic sentences using conjunction, disjunction and negation. We take each way of imagining a sentence true to be described by a set of literal facts; that is, facts corresponding to atomic sentences and their negations. We simply take it for granted that an agent presents some representational content to herself when imagining an atomic sentence true or false. We assume, moreover, that the representational content of atomic sentences is unstructured, in the sense that each literal (atomic sentence or negation thereof) corresponds to a single imaginative content. The logical operations of negation, disjunction and conjunction then take these contents of literals and combine them to form new contents.

To construct this semantic object, designed to represent the ways of imagining a given sentence true, we do not take a meaning to be given by the set of worlds at which it is true. An utterance such as 'Imagine you are drinking a mojito, reading a book on the beach,' is not an assertion but an imperative. As Frege put it, "Such a clause [an imperative] has no reference but only a sense" (Frege, 1948, 220). And as is well-known, imperatives pose a great challenge to accounts of meaning in terms of possible worlds (see Charlow 2014, though Lewis 1972, 1979 attempts to give such an account).

The present approach is instead indebted to developments of more fine-grained perspectives on meaning, such as situation semantics (Barwise and Perry 1983) and state-based approaches like inquisitive (Ciardelli et al. 2013) and truthmaker semantics (Fine 2017). Indeed, the present account draws more from these approaches than just a more fine-grained notion of meaning. The logic's disjunction corresponds to that of inquisitive semantics, and its conjunction to that of truthmaker semantics. And it is fair to say that, were it not for Kit Fine's recent work in exact semantics (e.g. Fine 2016), this essay would have been twice the length and said half as much.

1.1 Why a logic of imaginative content?

The main philosophical motivation for this essay is to better understand the meaning of counterfactuals, and in particular, counterfactual antecedents. The idea is this: *counterfactual antecedents evoke hypothetical contexts*. They prompt us to imagine scenarios and see whether the consequent is true there.

This in itself is quite a platitude. It is some clichè to point out that we need imagination to reason about counterfactuals—a departure from actuality is in their very name. (As Alison Gopnik once put it, "counterfactuals are the price we pay for hypotheticals".[1]) But it is one thing to acknowledge the imaginative character of counterfactuals, and quite another to make a semantic theory out of it. The role of imagination in evaluating the truth of counterfactuals may be unremarkable, but sometimes the stable, robust surface of a platitude is just what one needs to build a theory. And this is what we set out to do, motivated by the following approach.

> **The present approach to the meaning of counterfactuals.**
>
> To understand the semantic contribution of the words *and*, *or* and *not* in counterfactual antecedents, we must understand their contribution in evoking hypothetical contexts.

On a terminological note, we will often talk about a sentence's 'imaginative content'. What we mean by this is the interpretation the sentence receives when it appears in hypothetical contexts.

Now, it may rightfully be asked whether the present approach is anything new. One may even argue that possible worlds are themselves a kind of imaginative object, with each world a maximal hypothetical context. And from there, one could say that possible worlds semantics is the exercise of a hyperbolic imaginative act, by defining the meaning of *and*, *or* and *not* in terms of operations on these maximal hypothetical contexts (those operations being intersection, union and complementation, respectively). And so, it may be asked, if possible worlds semantics is already 'imaginative', why do we need a new approach, one with its own logic? Won't classical logic do?

The following section is devoted to showing why it will not.

[1] In personal communication with Huw Price and Brad Westlake (2009, 430).

1.2 The unclassical imagination

When evoking hypothetical contexts, the meanings of *and*, *or* and *not* differ in a number of striking ways from their truth-functional interpretation in classical logic. This section is intended as a tour through just some of the nonclassical features of conjunction, disjunction and negation in hypothetical contexts.

1.2.1 Disjunctions are not inclusive

I currently do not have a kitten or a puppy. However much I wish I did, the closest thing I have to a pet is the bacteria festering in my kitchen. But that does not stop my imagination. I often imagine that, if I had a kitten or a puppy, although I would have a little mouth to feed, I would come home to one little belly to rub and two little eyes looking up at me—one little face to photograph and post online. So keeping in mind that I actually have no pets at the moment, imagine I had a kitten or a puppy. Then how many pets would I have?

It seems I would have just one. Disjunction appears to not be inclusive in hypothetical contexts. Here are some examples in dialogue.

(2) [A *and* B *currently have no pets.*]

A: I think we could spend a maximum of €20 per week on pet food.
B: Well according to this website, on average, each cat costs €12 per week and each dog costs €15 per week.
A: So if we had a cat or a dog, we would be within our budget.
B: Yes, that's good news!

Things would sound false were we to explicitly force an inclusive reading of the disjunctive antecedent above.

(3) A: # If we had a cat or a dog, or both a cat and a dog, we would be within our budget.

And (4) is a variation on the theme of (1a) above.

(4) A: I've finished setting the table. You said five people, right?
B: I think you should set one more place, in case Mina or Rachael come.

Unless Mina and Rachael are willing to share, B would not want to say (5).

(5) B: #I think you should set one more place, in case Mina or Rachael or both come.

Let φ and ψ stand for arbitrary sentences built out of atomic sentences, negation, conjunction and disjunction. Then the lesson we draw from these examples is that a logic of imaginative content should not validate the equivalence of $\varphi \vee \psi$ and $\varphi \vee \psi \vee (\varphi \wedge \psi)$. Semantically speaking, this is to say that a logic aiming to represent hypothetical contexts should not assign the same interpretation to $\varphi \vee \psi$ as to $\varphi \vee \psi \vee (\varphi \wedge \psi)$.

1.2.2 Negated conjunctions are inclusive

While disjunctions are not inclusive in hypothetical contexts, it seems negated conjunctions are. We consider two examples. The first (6) is reminiscent of (1b).

(6) [SAÚL, JONATHAN *and* MORWENNA *are chatting at the wedding of their college friends, Rob and Nora.*]

SAÚL:	Oh what a fantastic day! You remember my party when Nora and Rob hit it off for the first time? That's where this all began.
JONATHAN:	I know. If they hadn't both come to your party that one night, we wouldn't be sitting here at their wedding today.
MORWENNA:	Hmm, I'm not so sure. Don't you remember? Saúl's party happened on the same night as the colloquium. If Robert and Noor had not both come to Saúl's party, then they might have both gone to the colloquium drinks and hit it off there instead.
JONATHAN:	Good point.

Compare what Morwenna said with what what she could have said, if negated conjunctions were not inclusive, meaning that *only one* of the conjuncts were false.

(7) MORWENNA: If Rob and Nora were not both at Saúl's party, then Rob would have been there, or Nora would have been there.

It may be factually correct that at least one of Rob and Nora were always going to attend the party. But at least one of them going does not follow as a matter of logic from the hypothetical, 'If Rob and Nora had not both gone, ...'.

The second example comes from Ciardelli, Zhang, and Champollion (2018b). Consider two switches connected to a light bulb, wired as in Figure 1.1.

As Figure 1.1 depicts, the light is on just in case both switches are in the same position (i.e. both up or both down). If the switches are in different positions, the light is off. In the actual situation, both switches are up. Then what are the truth values of the following counterfactuals?

(8) a. If switch A or switch B was down, the light would be off.
    b. If switch A and switch B were not both up, the light would be off.

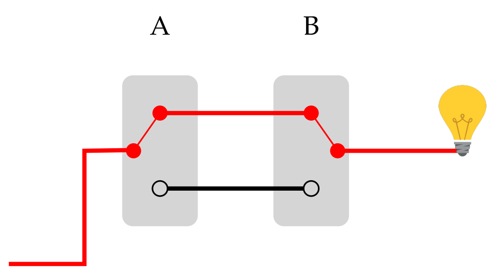

Figure 1.1: Two switches

Comparing 1425 responses, Ciardelli et al. (2018b) found a significant difference between the two sentences. (8a) was judged true by a wide majority, whereas (8b) was generally judged false or indeterminate. The authors conclude:

> it seems that most participants interpreted [(8a)] by considering one switch at a time, while ignoring the option that both switches might be toggled simultaneously.
>
> ...the predominant strategy for [(8b)] is to consider all three possibilities: only switch A is toggled; only switch B is toggled; both switches are toggled. These possibilities do not all agree on the state of the light, leading to the lack of 'true' judgments.

Ciardelli et al. (2018b) draw the moral from their experiment that De Morgan's first law, stating the equivalence of $\neg(\varphi \wedge \psi)$ and $\neg\varphi \vee \neg\psi$, fails when it comes to counterfactual antecedents. To put things in terms of the present account, the moral we draw is that a logic of imaginative equivalence should not validate De Morgan's first law; that is, the equivalence of $\neg(\varphi \wedge \psi)$ and $\neg\varphi \vee \neg\psi$.

1.2.3 Superfluous information changes meaning

Recalling (1c) above, compare the following.

(9) a. I put milk, or milk and hemlock in your tea.
b. I put milk in your tea.

On a truth-conditional picture, the meaning of the logical connectives can be described in terms of operations on *information*. When it comes to a disjunction's informative content, say, it makes sense to say that a disjunct contains the information of any disjunction in which it occurs.[2] For instance, the information that it I put milk and helmock in your tea contains the information that I

[2]By a sentence's 'informative content' we mean simply its truth conditions; or, with respect to a given model, the set of worlds at which it is true.

put milk in your tea. And so we might say (9a) contains a superfluous disjunct. Thinking purely in terms of information, as classical logic does, (9a) and (9b) are equivalent, in the sense that the information of each is contained in that of the other. But that does not stop their meaning from pulling apart in hypothetical contexts such as (10) and (11).[3]

(10) a. Imagine I put milk, or milk and hemlock in your tea.
b. Imagine I put milk in your tea.

(11) a. Had I put milk, or milk and poison in your tea, you would have enjoyed it.
b. Had I put milk in your tea, you would have enjoyed it.

Clearly, (10a) and (10b) raise different hypothetical contexts. So a logic of imaginative equivalence ought to distinguish them: a logic of imaginative content should not validate the equivalence of φ and $\varphi \vee (\varphi \wedge \psi)$.

1.2.4 Disjunction does not distribute over conjunction

To say that disjunction distributes over conjunction is to say that $\varphi \vee (\psi \wedge \chi)$ is equivalent to $(\varphi \vee \psi) \wedge (\varphi \vee \chi)$. This equivalence does hold in classical logic, though the following example is designed to show that it should not hold in a logic of imaginative equivalence. We constructed quite a long—and contrived!—dialogue in order to get a sentence of the form $(\varphi \vee \psi) \wedge (\varphi \vee \chi)$ to even be assertable.

(12) [*The results of a local election have just been announced. The three candidates who did not secure a seat were Beth, Parker and Stuart. Beth is both a student and a parent. Parker is a parent, and Stuart is a student.*]

A: I thought Beth, Parker and Stuart gave great speeches.

B: It's a shame the new committee doesn't have any students or parents on it. It's very unrepresentative. If Beth, or Parker and Stuart had been elected, I would be happier with the result.

A: Hold on, why should the committee need both Parker *and* Stuart? I thought you meant that you would be happier if the committee had featured a student and a parent. Then wouldn't you also be happy with the result if, say, Beth and Parker had been elected? After all, Parker is a parent, and Beth is both a student and a parent.

B: Right. I suppose I didn't mean to imply I'd only be happier if Beth or both Parker and Stuart had been elected. Beth and Parker are parents, and Beth and Stuart are students. I guess what I really

[3] Thanks to Robert van Rooij, in whose philosophical logic class I first encountered the problem of distinguishing sentences such as (10a) and (10b).

meant to say was that if the committee had featured Beth or Parker, and Beth or Stuart, I would be happier.

A: I agree. If Beth or Parker, and Beth or Stuart had been elected, the committee would be more representative.

In the dialogue, B originally imagines being happier with the result if Beth, or Parker and Stuart had been elected. Then A disputes that this is the only way for B to be happier with the result, arguing that they would also be happier if Beth or Parker, and Beth or Stuart had been elected. A's challenge only makes sense if (13a) and (13b) are not equivalent in hypothetical contexts.

(13) a. The committee elected Beth, or Parker and Stuart.
b. The committee elected Beth or Parker, and Beth or Stuart.

Clearly, A has a point, making a coherent objection to B's original assertion. And consider (14) in a context where Beth and Stuart hate one another, refusing to work together.

(14) a. If Beth, or Parker and Stuart had been elected, we would have had a happy committee.
b. # If Beth or Parker, and Beth or Stuart had been elected, we would have had a happy committee.

What we draw from this example, then, is that a logic of imaginative equivalence should not validate the distribution of disjunction over conjunction.

1.3 Equivalences in three counterfactual semantics

Equivalence is, semantically speaking, a stubborn notion. For two sentences to be semantically equivalent is for each to be assigned an identical interpretation. This ensures that if a semantics makes two sentences equivalent, by the strength of the identity relation there is no possible semantic operation one could define over their interpretations to pull them apart again. This is why we focused on equivalences in the above discussion. As expressions of identity, equivalences showcase a semantics at its most dogmatic.

The counterfactual semantics beginning with Stalnaker (1968) and Lewis (1973b) interpret counterfactual antecedents over possible worlds; that is to say, classically. It is also worth comparing the equivalences of two other systems, inquisitive semantics and truthmaker semantics. We look at these two systems because they have recently been applied to counterfactual semantics: inquisitive semantics in Ciardelli et al. (2018b); truthmaker semantics in Fine (2012a). Table 1.1 compares what equivalences are *invalid* in the respective approaches.

In light of the previous section, if the semantics in question is to successfully apply to hypothetical contexts, a tick in the table is good sign.

CAN THE EQUIVALENCE FAIL?	Possible worlds semantics	Inquisitive semantics	Truthmaker semantics
Inclusive disjunction $\varphi \vee \psi \Leftrightarrow \varphi \vee \psi \vee (\varphi \wedge \psi)$	✗	✗	✗ / ✓
De Morgan's first law $\neg(\varphi \wedge \psi) \Leftrightarrow \neg\varphi \vee \neg\psi$	✗	✓	✗
Adding superfluous information $\varphi \Leftrightarrow \varphi \vee (\varphi \wedge \psi)$	✗	✗	✓
Distribution of $\vee$ *over* $\wedge$ $\varphi \vee (\psi \wedge \chi) \Leftrightarrow (\varphi \vee \psi) \wedge (\varphi \vee \chi)$	✗	✗	✓

Table 1.1: Comparing invalid equivalences in three counterfactual semantics

The table shows that each approach validates some equivalence we saw (1.2) a semantics of counterfactual antecedents should not validate. To bring the problem into relief, we have repeated some examples below. Any counterfactual semantics based on possible world semantics assigns each sentence of (15)–(18) the same truth value. Likewise, each sentence of (15), (17) and (18) has the same truth condition according to any counterfactual semantics using inquisitive propositions; similarly for (16) and truthmaker semantics.

(15) [*The budget is €20 per week. A cat is €12 per week, a dog €15 per week.*]
  a. If we had a cat or a dog, we would be within our budget.
  b. If we had a cat or a dog or both, we would be within our budget.

(16) [*Switches A and B are currently both up.*]
  a. If switch A or switch B were down, the light would be off.
  b. If switch A and switch B were not both up, the light would be off.

(17) a. You would drink the tea, had I put milk in it.
  b. You would drink the tea, had I put milk, or milk and hemlock in it.

(18) [*Beth and Stuart refuse to work together.*]
  a. If Beth, or Parker and Stuart had been elected, we would have had a happy committee.
  b. If Beth or Parker, and Beth or Stuart had been elected, we would have had a happy committee.

To show that inquisitive semantics does not validate De Morgan's first law, let $[\![\varphi]\!]$ be the proposition that inquisitive semantics assigns to a sentence φ. Then in inquisitive semantics (Ciardelli et al., 2018a), the information state of $\varphi \vee \psi$ being true supports the proposition $[\![\neg(\varphi \wedge \psi)]\!]$ but not the proposition $[\![\varphi \vee \psi]\!]$, which shows that $\neg(\varphi \wedge \psi)$ and $\varphi \vee \psi$ are not inquisitively equivalent.

The reason for the cross and tick for inclusive disjunction in truthmaker semantics in Table 1.1 is that Fine (2012a) presents both inclusive and exclusive variants of truthmaker semantics. Fine himself gives his semantics for counterfactuals in terms of an *inclusive* version, which does validate inclusive disjunction. Nonetheless, he does allow for one to instead take an exclusive version of truthmaking semantics, where inclusive disjunction is invalid. (The remaining equivalences of the table do not depend on whether one takes the inclusive or exclusive variants.) In section A1 of the appendix we show that De Morgan's first law is valid in both the inclusive (Fact 52) and exclusive (Fact 53) variants of truthmaker semantics.

The narrative arc of this essay is the slow and steady working of the notion of imaginative content into a semantics of counterfactuals. But instead of treating imagination from prior notions such as truth (as in possible worlds semantics), the meaning of questions (inquisitive semantics) or exact verifiers for sentences (truthmaker semantics), we will treat the logic of imagination on its own terms. For, the presence of at least one cross in Table 1.1 with respect to each counterfactual semantics above signals the need for a new approach. With the semantic stage thus set, let us now aim to capture the logical behaviour of imaginative content.

Chapter 2

A logic of imaginative equivalence

In the previous section we saw that *and*, *or* and *not* take on a unique meaning when they are used to evoke hypothetical contexts. This is not the same as their meaning in making assertions, described by possible worlds semantics, nor it is the same as their meaning according to inquisitive or truthmaker semantics. Now, that negative answer is not wholly satisfying. We have to wonder: What *do* the words 'and', 'or' and 'not' mean when used to evoke hypothetical contexts?

We will approach the question from an inferential as well as semantic point of view, describing not just the semantics but the logic of imaginative content, which includes a deductive system. I can think of three reasons to care about deductive systems. Firstly, an axiomatisation offers a more well rounded perspective on the semantics; and as we will see, in this case also offers a much clearer view of the logic than the semantic clauses themselves provide. Secondly, by having a logic of imaginative content for which to prove soundness and completeness, the logic can be embedded into more familiar logics—such as propositional and modal logics—and we may then prove soundness and completeness for those logics enriched with a notion of imaginative content. Lastly, an axiomatisation exposes the inferential role of *and*, *or* and *not* as they appear in imaginative contexts, allowing us to capture to meaning of those connectives according to an inferentialist picture of meaning (e.g. Brandom, 2000).

The task is to create a logic whose conjunction, disjunction and negation correspond as closely as possible to *and*, *or* and not, as they are used to evoke hypothetical contexts. Now, there is one issue to be worked out before this can happen. We aim to analyse the meaning of these three connectives as they appear, not in assertions, but in invocations to imagine such as, "Imagine eating coconut ice cream". But how exactly can a logic—as language, deductive system and semantics—ask us to imagine anything? A logic asserts its theorems. Yet here we are trying to get away from the meaning of *and*, *or* and *not* as they appear in assertions. And so our problem is to create a logic whose assertions capture the meaning of words that is unlike their meaning in assertions.

Our solution to this problem rests on the observation that, although we cannot assert imaginative content itself, we can assert when two sentences have the same imaginative content. The trick is to not create a logic of imaginative content directly but a logic of imaginative *equivalence*. This logic, called IE, will articulate when two sentences are 'imaginatively equivalent' in the sense that the ways of imagining one are all and only the ways of imagining the other. This logic will then serve as a foundation when constructing further relationships between imaginative content.

Let us now turn to the logic of imaginative equivalence.

2.1 Language of IE

We first define a *base language* $\mathcal{L}$, whose sentences are built up from atomic sentences, negation, conjunction and disjunction. That is, every sentence φ of $\mathcal{L}$ is one of the following forms, where p is any atomic sentence and $\psi, \chi \in \mathcal{L}$.

$$\varphi ::= p \mid \neg\psi \mid \psi \wedge \chi \mid \psi \vee \chi$$

Then let *equivalential language* of $\mathcal{L}$, denoted $\mathcal{L}_{\equiv}$, consists of equivalences of sentences in $\mathcal{L}$. That is, we take $\mathcal{L}_{\equiv}$ to be $\{\varphi \equiv \psi : \varphi, \psi \in \mathcal{L}\}$. This equivalential language $\mathcal{L}_{\equiv}$ will serve as the language of IE.

2.2 Axiomatisation of IE

In this section we provide an axiomatisation for the logic of imaginative equivalence and state some basic facts of the deductive system. Where p is any atomic sentence, and l any literal, the axioms and rule of inference of IE are all instances of the following schema.

Axiom schema of IE

Atomic double negation	$p \equiv \neg\neg p$	(A1)
Literal $\wedge$ idempotence	$l \equiv l \wedge l$	(A2)
$\vee$ idempotence	$\varphi \equiv \varphi \vee \varphi$	(A3)
Communitativity	$\varphi \vee \psi \equiv \psi \vee \varphi$	(A4)
	$\varphi \wedge \psi \equiv \psi \wedge \varphi$	(A5)
Associativity	$\varphi \vee (\psi \vee \chi) \equiv (\varphi \vee \psi) \vee \chi$	(A6)
	$\varphi \wedge (\psi \wedge \chi) \equiv (\varphi \wedge \psi) \wedge \chi$	(A7)
Distribution of $\wedge$ over $\vee$	$\varphi \wedge (\psi \vee \chi) \equiv (\varphi \wedge \psi) \vee (\varphi \wedge \chi)$	(A8)
Inclusive De Morgan I	$\neg(\varphi \wedge \psi) \equiv (\neg\varphi \vee \neg\psi) \vee (\neg\varphi \wedge \neg\psi)$	(A9)
De Morgan II	$\neg(\varphi \vee \psi) \equiv \neg\varphi \wedge \neg\psi$	(A10)
Negation elimination	$\neg\varphi \equiv \neg\neg\neg\varphi$	(A11)
Distribution of $\neg\neg$ over $\wedge$	$\neg\neg(\varphi \wedge \psi) \equiv \neg\neg\varphi \wedge \neg\neg\psi$	(A12)
$\neg\neg$ De Morgan	$\neg\neg(\varphi \vee \psi) \equiv \neg(\neg\varphi \wedge \neg\psi)$	(A13)

Rules of inference of IE

$$\frac{\varphi \equiv \psi}{\psi \equiv \varphi}\ \text{Symmetry} \qquad \frac{\varphi \equiv \psi \qquad \psi \equiv \chi}{\varphi \equiv \chi}\ \text{Transitivity}$$

$$\frac{\varphi \equiv \psi}{\varphi \wedge \chi \equiv \psi \wedge \chi}\ \wedge\ \text{Addition} \qquad \frac{\varphi \equiv \psi}{\varphi \vee \chi \equiv \psi \vee \chi}\ \vee\ \text{Addition}$$

Reflexivity of $\equiv$ follows straightforwardly from symmetry and transitivity.

Fact 1 (Reflexivity)**.** *$\varphi \equiv \varphi$ is a theorem of IE.*

Proof.

$$\dfrac{\dfrac{}{\varphi \equiv \varphi \vee \varphi}\ \text{A3} \qquad \dfrac{\dfrac{}{\varphi \equiv \varphi \vee \varphi}\ \text{A3}}{\varphi \vee \varphi \equiv \varphi}\ \text{Symmetry}}{\varphi \equiv \varphi}\ \text{Transitivity}$$

□

Thus, $\equiv$ is an equivalence relation, and so merits the name 'imaginative equivalence'.

We turn now to the interpretation of IE.

2.3 Interpretation of IE

The present approach to hypothetical contexts is guided by the idea that when we imagine a sentence true, we represent some scenario to ourselves. This scenario may be partial—it may even be inconsistent. All we require for a formal account of such acts of imagination is that the imagined scenario be representable by a list of literal sentences. Certainly, imagining agents do not have to be logically omniscient, but for their imagination to admit a formal treatment we assume that every detail of their imagined scenario can given a name; that is to say, corresponds to a literal sentence. For without such an Edenic assumption, imagined scenarios could hardly be communicated, neither to oneself nor others, and so there would be no language for a logic of imaginative content to speak.

Now, there may be many ways to imagine a sentence true. If asked to imagine being at home or at work, it seems there are two ways to do it: to imagine being at home and to imagine being at work. And if asked to imagine it false that two given switches are both up, it seems there are three ways to do that: imagine one switch down, or the other down, or both down.

This tells us how much structure we need to build an interpretation function for imaginative content. We take it that the imaginative content of a sentence is given by a set of sets. Literal sentences have a particularly simple structure. Since there is only one way to imagine an atomic sentence true or false, the imaginative content of a literal is a singleton. To keep our semantic account sufficiently general, we will not specify directly what it means to imagine a literal sentence true: it may involve some sensory or verbal content. All we require is that to each literal sentence there corresponds *some* basic content. And so we define a model in the following manner.

Definition 2 (Model of IE). *Fix now and henceforth a set X (of "contents"). A model of IE is a pair (X, ι) where $\iota : l \mapsto x$ assigns to each literal l an element x of X.*

The set of basic contents X may contain images, sounds, ideas, such stuff as dreams are made on—whatever. What exactly it means to imagine an atomic sentence true or false is a question the logic need not settle, at least not at the formal stage on which this essay plays, lacking input from psychology and cognitive science. Let us briefly elaborate the reasons for including this set X of contents in our interpretation.

2.3.1 Abstract and representational semantics

Since our semantics assumes a set of so-called 'cognitive contents', we might call it a *representational* semantics. But we could also give an *abstract* semantics, let's call it, which is abstract in the sense that such talk of content does not appear at all. To see this, note that, in effect, what the above definition of models of IE provides is a partition over the set of literals, specifying whether literals are assigned the same or different content. Where $\mathbb{L}$ is the set of literals, then, given a model of IE we may generate its corresponding equivalence relation $\sim \subseteq \mathbb{L} \times \mathbb{L}$ defined so that $l \sim l'$ just in case $\iota(l) = \iota(l')$, for any literals $l, l' \in \mathbb{L}$. From this abstraction, we can take the *abstracted content* of a literal l to be its equivalence class $[l] = \{l' \in \mathbb{L} : l \sim l'\}$. Where $[\mathbb{L}] = \{[l] : l \in \mathbb{L}\}$ is the set of such equivalence classes, and $[\iota] : l \mapsto [l]$ assigns to each literal its equivalence class, what we have ended up is a model $([\mathbb{L}], [\iota])$ of IE in which all talk of cognitive content is abstracted away. The residue is a purely logical representation, which, unlike our talk of 'content', carries no reference to the mind or world. In this respect the abstracted semantics is reminiscent of the semantics of propositional logic, which does not talk about what it is for an atomic sentence to be true, but only whether or not it is true.

When it comes to giving the semantic behaviour of *and*, *or* and *not* in hypothetical contexts, it does not make a difference whether one uses the abstract or representational semantics. Either way we can specify how the imaginative contents $\iota(l)$ combine under the operations of negation, conjunction and disjunction. We will nonetheless stick to the representational definition of model in Definition 2, for two reasons. The first is that a representational semantics is more general (for as we just saw, every representational model induces an abstract one; but one cannot recover contents from mere equivalences of literals). The second reason is that a representational semantics respects the idea that when we imagine a sentence true, we imagine *something*. That is to say, the imagination can invoke representations, rather than just the austere consideration of what other sentences are equivalent to the one we are asked to imagine.

The deeper concern of these remarks is that, under an abstract semantics, one could be tricked into believing that the task of describing the behaviour of imaginative content is a task for logicians alone. It is all too easy to forget what one takes for granted; in this case, the representational character of our imagination—a fact of human psychology. The awkward presence of a mystery set X of contents acts in the semantics as a reminder that a logical treatment of imagination should not occur in isolation from the work of psychology and cognitive science on which it ultimately relies.

2.3.2 Extending the interpretation function

To extend the imaginative content function i to arbitrary sentences, we make use of bivalent clauses. To each sentence φ of the base language $\mathcal{L}$ we associate a positive content $i(\varphi)$ and a negative content $i^-(\varphi)$. The idea is that $i(\varphi)$ contains the ways of imagining φ true, while $i^-(\varphi)$ contains the ways of imagining the sentence false.

Positive conjunction and negative disjunction are given by an operation of *pairwise union*. For any sets A and B, we define their pairwise union like so.

$$A \Cup B = \{a \cup b : a \in A, b \in B\}$$

For instance, if $A = \{\{a\}\}$ and $B = \{\{b\}, \{c\}\}$, then $A \Cup B = \{\{a, b\}, \{a, c\}\}$. The other operation we use is *closure under union*. Given a set A, we define its closure under union, denoted $A^{\cup}$, as follows.

$$A^{\cup} = \{\bigcup A' : A' \subseteq A \text{ and } A' \neq \emptyset\}$$

For instance, if $B = \{\{b\}, \{c\}\}$ then $B^{\cup} = \{\{b\}, \{c\}, \{b, c\}\}$.

We then extend the imaginative content function to arbitrary sentences of $\mathcal{L}$ as follows.

Definition 3 (Interpretation of IE)**.** *The imaginative content of a sentence φ of $\mathcal{L}$ is a pair $i^{\pm}(\varphi) = (i(\varphi), i^{-}(\varphi))$ containing a positive and negative content, where*

$$\begin{aligned}
i(p) &= \{\{\iota(p)\}\} & i^{-}(p) &= \{\{\iota(\neg p)\}\} \\
i(\neg\varphi) &= i^{-}(\varphi)^{\cup} & i^{-}(\neg\varphi) &= i(\varphi) \\
i(\varphi \wedge \psi) &= i(\varphi) \Cup i(\psi) & i^{-}(\varphi \wedge \psi) &= i^{-}(\varphi) \cup i^{-}(\psi) \\
i(\varphi \vee \psi) &= i(\varphi) \cup i(\psi) & i^{-}(\varphi \vee \psi) &= i^{-}(\varphi) \Cup i^{-}(\psi)
\end{aligned}$$

How do these clauses sound in English? Given a model (X, ι) of IE and a set $C \subseteq X$ of contents, they amount to the following.

- $i(l)$ C is a way of imagining a literal l true if and only if (iff) $C = \{\iota(l)\}$.
- $i(\neg\varphi)$ C is a way of imagining $\neg\varphi$ true iff it is a union of ways of imagining φ false.
- $i^{-}(\neg\varphi)$ C is a way of imagining $\neg\varphi$ false iff it is a way of imagining φ true.
- $i(\varphi \wedge \psi)$ C is a way of imagining $\varphi \wedge \psi$ true iff it is the union of a way of imagining φ true and a way of imagining ψ true.
- $i^{-}(\varphi \wedge \psi)$ C is a way of imagining $\varphi \wedge \psi$ false iff it is a way of imagining φ false or a way of imagining ψ false.
- $i(\varphi \vee \psi)$ C is a way of imagining $\varphi \vee \psi$ true iff it is a way of imagining φ true or a way of imagining ψ true.
- $i^{-}(\varphi \vee \psi)$ C is a way of imagining $\varphi \vee \psi$ false iff it is the union of a way of imagining φ false and a way of imagining ψ false.

Three points about the clauses are in order. Firstly, it is worth clarifying that when we speak of a logic's 'interpretation' or 'semantics', we do so in the logician's sense as simply an assignment of mathematical objects to sentences. We do not wish to claim to be offering a semantic explanation—in the linguist's sense—of why φ receives the interpretation that it does. For example, IE's disjunction is not inclusive: $i(\varphi \vee \psi)$ is generally distinct from $i(\varphi \vee \psi \vee (\varphi \wedge \psi))$. This is not meant to imply the problematic view (rejected by Horn, 1985; Grice, 1991) that *or* is lexically ambiguous between inclusive and noninclusive meanings. It may well be that *or*'s inclusive meaning is blocked in hypothetical contexts by pragmatic factors, in which case IE's "semantics" would be ultimately explained by recourse to semantic and pragmatic theory.[1]

Secondly, making $i(l)$ a singleton for any literal l amounts to the claim that there is only one way to imagine a literal sentence true. The idea here is that imagining a literal true is a particularly simple cognitive act, one comprised of a single content.

Thirdly, by taking the ways of imagining $\neg\varphi$ true to be unions of ways of imagining φ false, we take the imagining of negations to be sloppy, so to speak, in that one may imagine more than is strictly necessary to imagine the sentence false. We saw an illustration of this phenomenon in section 1.2.2 in the light switch example of Ciardelli et al. (2018b). There we had two binary variables representing the positions of switches A and B, and so four specifications of their positions in all. When asked to imagine one specification false (say, imagine both switches were not both up), participants seems to imagine three specifications. We might conjecture—though without further evidence—that the participants considered the four possible specifications of the switches and, when asked to imagine the switches not both up, simply removed that scenario from consideration of the four. Of the three remaining, one specification even featured both switches down, which is strictly more than one needs for the switches to not both be up.

In more technical terms, by taking the ways of imagining $\neg\varphi$ true to be unions of ways of imagining φ false we can account for the asymmetry between conjunction and disjunction, whereby disjunctions are not inclusive but negated conjunctions are.

Moving on, the definition of truth in a model should not come as a surprise.

Definition 4 (Truth). *$\varphi \equiv \psi$ is true a model (X, ι) of IE just in case $i(\varphi) = i(\psi)$.*

In other words, $\varphi \equiv \psi$ is true in a model just in case, according to it, φ and ψ are "imaginatively equivalent": the ways of imagining φ true are all and only ways of imagining ψ true.

[1]For more on linguistic explanations of the behaviour of hypothetical contexts, see section 6.3.

What is a surprise, however, is that even though we are working with a fine-grained semantics, with many contents in place of truth values, our logic is still truth-conditional. The logic only talks about sentences having the same imaginative content, and not talking directly about those imaginative contents themselves.[2] Despite the fine-grained semantic objects, then, this essay will stay by Lewis's point, made almost half a century ago, that "A meaning for a sentence is something that determines the conditions under which the sentence is true or false" and, more emphatically, that "Semantics with no treatment of truth conditions is not semantics" (Lewis 1972). The larger point is that logics of equivalence offer a useful bridge between fine-grained and truth-conditional approaches to meaning.

2.4 Soundness

The deductive system of IE is sound respect to the semantics of Definition 3 above. Since the proof of soundness is relatively technical and routine, we leave the details to the appendix.

Theorem 5 (Soundness)**.** *Every theorem of IE is true in every model of IE.*

Proof. By set-theoretic argument. Theorem 59 of the appendix. □

2.5 Some select invalidites of IE

IE is quite different from classical logic. As an example, $(p \vee q) \wedge (p \vee q)$ is equivalent to $(p \vee q) \vee (p \wedge q)$, rather than simply $p \vee q$. This is because, from the perspective of IE, every time we to imagine a conjunction of disjunctions, we may pick one disjunct from each conjunct. So when asked to imagine that Alice or Bob attended the party, we may pick Alice, and when asked *again* to imagine Alice or Bill attended the party, we may pick Bill: putting the two choices together, we may imagine that Alice and Bill both attended the party.

Rather than discussing all notable invalidities of IE on a case by case basis, we will enumerate some and give a countermodel to each. Notice that (2), (3), (4) and (6) below are the equivalences that in the introduction (1.2) we saw a logic of imaginative content ought not to validate.

[2] Another equivalential logic is Fine's axiomatisation of Angell's logic AC (Fine, 2016, 201).

Proposition 6 (Invalidities)**.** *The following are not validities of IE. (For each equivalence there are φ, ψ, χ of $\mathcal{L}$ and a model of IE in which the equivalence does not hold.)*

1.	$\varphi \equiv \neg\neg\varphi$	(Double negation elimination)
2.	$\neg(\varphi \wedge \psi) \equiv \neg\varphi \vee \neg\psi$	(De Morgan's first law)
3.	$\varphi \vee \psi \equiv (\varphi \vee \psi) \vee (\varphi \wedge \psi)$	(Inclusive disjunction)
4.	$\varphi \vee (\psi \wedge \chi) \equiv (\varphi \vee \psi) \wedge (\varphi \vee \chi)$	(Distribution of $\vee$ over $\wedge$)
5.	$\varphi \equiv \varphi \wedge \varphi$	(Idempotence of $\wedge$)
6.	$\varphi \equiv \varphi \vee (\varphi \wedge \psi)$	(Disjoining stronger info.)
7.	$\varphi \equiv \varphi \wedge (\varphi \vee \psi)$	(Conjoining weaker info.)
8.	$\varphi \vee \neg\varphi \equiv \psi \vee \neg\psi$	(Uniqueness of $\top$)
9.	$\varphi \wedge \neg\varphi \equiv \psi \wedge \neg\psi$	(Uniqueness of $\bot$)
10.	$\neg(\varphi \wedge (\psi \vee \chi)) \equiv \neg((\varphi \wedge \psi) \vee (\varphi \wedge \chi))$	

Hence, by soundness, the above equivalences are not theorems of IE either.

Proof. We need only consider the model $(\mathbb{L}, \mathrm{id}_{\mathbb{L}})$ where $\mathbb{L}$ is the set of literals and $\mathrm{id}_{\mathbb{L}} : l \mapsto l$ the identity map. (By distinguishing all literals, $(\mathbb{L}, \mathrm{id}_{\mathbb{L}})$ validates as few equivalences as possible.) The single counterexample of $\{\{p\},\{q\}\} \neq \{\{p\},\{q\},\{p,q\}\}$ suffices to refute (1–3) and (5).

$$\begin{aligned} i(p \vee q) = i(\neg\neg p \vee \neg\neg q) &= \{\{p\},\{q\}\} \\ &\neq \{\{p\},\{q\},\{p,q\}\} \\ &= i(\neg\neg(p \vee q)) && (1) \\ &= i(\neg(\neg p \wedge \neg q)) && (2) \\ &= i((p \vee q) \vee (p \wedge q)) && (3) \\ &= i((p \vee q) \wedge (p \vee q)) && (5) \end{aligned}$$

And for the remaining clauses we calculate as follows.

$$\begin{aligned} i(p \vee (q \wedge r)) &= \{\{p\},\{q,r\}\} \neq \{\{p\},\{p,q\},\{p,r\},\{q,r\}\} \\ &= (p \vee q) \wedge (p \vee r) && (4) \\ i(p) &= \{\{p\}\} \neq \{\{p\},\{p,q\}\} = i(p \vee (p \wedge q)) && (6) \\ &= i(p \wedge (p \vee q)) && (7) \\ i(p \vee \neg p) &= \{\{p\},\{\neg p\}\} \neq \{\{q\},\{\neg q\}\} = i(q \vee q) && (8) \\ i(p \wedge \neg p) &= \{\{p, \neg p\}\} \neq \{\{q, \neg q\}\} = i(q \wedge \neg q) && (9) \\ i(\neg(p \wedge (q \vee r))) &= \{\{\neg p\},\{\neg q, \neg r\},\{\neg p, \neg q, \neg r\}\} \\ &\neq \{\{\neg p\},\{\neg p, \neg q\},\{\neg p, \neg r\},\{\neg q, \neg r\},\{\neg p, \neg q, \neg r\}\} \\ &= i(\neg((p \wedge q) \vee (p \wedge r))) && (10) \end{aligned}$$

$\square$

Let us now quickly pause to consider some invalidities of IE that are of special theoretical semantic interest. This time they concern disjunction, and in particular, whether it can be called 'inclusive' or 'exclusive'.

2.5.1 The false inclusive/exclusive dichotomy

What does it mean for disjunction to be interpreted 'inclusively' or 'exclusively'? The following definitions seem natural, where we let $[\![\varphi]\!]$ be the interpretation a sentence φ receives under the semantics in question.

$$[\![\varphi \vee \psi]\!] = [\![\varphi \vee \psi \vee (\varphi \wedge \psi)]\!] \qquad \text{(Definition of inclusive } \vee\text{)}$$
$$[\![\varphi \vee \psi]\!] = [\![(\varphi \wedge \neg\psi) \vee (\psi \wedge \neg\varphi)]\!] \qquad \text{(Definition of exclusive } \vee\text{)}$$

To be precise we will say a semantics interprets disjunction inclusively or exclusively just in case it satisfies the respective equation above; in other words, disjunction is inclusive when "φ or ψ" semantically means "φ or ψ or both", and exclusive when it semantically means "φ but not ψ, or ψ but not φ".

Now for a surprising result: IE's disjunction is neither inclusive nor exclusive. According to our semantics of imaginative content, each of the following sentences receives a distinct semantic interpretation (in a model where for simplicity we put $\iota(l) = l$ for each literal l).

$$\begin{aligned} i(p \vee q) &= \{\{p\}, \{q\}\} \\ i(p \vee q \vee (p \wedge q)) &= \{\{p\}, \{q\}, \{p, q\}\} \\ i((p \wedge \neg q) \vee (q \wedge \neg p)) &= \{\{p, \neg q\}, \{q, \neg p\}\} \end{aligned}$$

So by IE's soundness (Theorem 5), since we have found a countermodel, IE proves neither $\varphi \vee \psi \equiv \varphi \vee \psi \vee (\varphi \wedge \psi)$ nor $\varphi \vee \psi \equiv (\varphi \wedge \neg\psi) \vee (\psi \wedge \neg\varphi)$. The upshot is that the contraries 'inclusive' and 'exclusive' are not always contradictories. Some disjunctions are neither inclusive nor exclusive.[3]

2.6 Substitution in IE

Given any sentences φ, ψ, χ of $\mathcal{L}$, let $\varphi[{}^{\chi}/_{\psi}]$ be the result of replacing every occurrence of χ in φ with ψ. It is worth pointing out that the following rule of IE substitution is not an admissible rule of IE.

[3]Fine's exclusive truthmaker semantics is another framework whose disjunction, according to our definitions, is neither inclusive nor exclusive (see the appendix, section A1). On the other hand, the disjunctions of classical, intuitionistic and inquisitive logic are all inclusive and not exclusive.

$$\frac{\psi \equiv \chi}{\varphi \equiv \varphi[{}^{\chi}/{}_{\psi}]}\ \text{IE substitution}$$

To illustrate, compare the theorem $p \equiv \neg\neg p$ of IE with $p \vee q \equiv \neg\neg(p \vee q)$, which, as we have just seen (Proposition 6), is not a theorem of IE.

It is an interesting task to pinpoint the exact source of the failure of IE substitution. This is in general difficult to do, though the following proposition goes some way toward an answer.[4]

Proposition 7. *Unnegated equivalents are salva aequivalente substitutable. That is,*

$$\frac{\psi \equiv \chi}{\varphi \equiv \varphi[{}^{\chi}/{}_{\psi}]}$$

is a derived rule of IE, where ψ does not occur under the scope of a negation in φ.

Proof. Appendix Fact 54. The proof relies essentially on $\vee$ and $\wedge$ addition. □

2.7 Completeness

We prove completeness by means of disjunctive normal forms. The proof uses the same techniques as Fine's proof that Angell's logic AC is complete with respect to truthmaker semantics (Fine 2016). The idea—there as well as here—is to put sentences into a provably equivalent form corresponding to their semantic representation. In the present context, by taking as our canonical model a model that makes true as few equivalences as possible, we show firstly that a sentence is a theorem of IE just in case it is true in the canonical model. Secondly, we show that the canonical model brings together all other models in the sense that an imaginative equivalence is true in the canonical model of IE just in case it is true in every model of IE. The fact that every validity of IE is a theorem of IE will follow as an easy corollary.

2.7.1 Disjunctive normal forms

Let us say that a sentence is in *conjunctive normal form* just in case it is a literal or a conjunction of literals, and a sentence is in *disjunctive normal form* just in case it is a conjunctive normal form or a disjunction of conjunctive normal forms. That is, a sentence φ of $\mathcal{L}$ is in disjunctive normal form just in case for some finite index I and set of finite indices $\{J_i\}_{i \in I}$, where each l_{ij} is a literal,

$$\varphi = \bigvee_{i \in I} \bigwedge_{j \in J_i} l_{ij} = (l_{0,0} \wedge \cdots \wedge l_{0,k}) \vee \cdots \vee (l_{n,0} \wedge \cdots \wedge l_{n,m}).$$

[4]To be exact, we say ψ *occurs under the scope of a negation in* φ just in case ψ is a subformula of χ for some subformula $\neg\chi$ of φ.

Let us fix some list of the literals of our language, furnishing a linear order over the literals. And let us say that a conjunctive normal form is *standard* just in case each conjunct appears according to the fixed order, with no repeats and with association to the right. This definition guarantees that any conjunctive normal forms built from the same literals as conjuncts will be identical.

Let us also fix a linear order over the standard conjunctive normal forms and say that a disjunctive normal form is *standard* just in case each disjunct appears according to the fixed order without repeats and with association to the right. This definition then guarantees that any sentences built from the same standard conjunctive normal forms as disjuncts will be identical.

Theorem 8. *Any sentence is provably equivalent to one in disjunctive normal form.*

Proof. In addition to the rules of inference, use (A10–13) to turn arbitrary negations into single negations, use De Morgan laws (A9–10) to push single negations inside conjunctions and disjunctions, and use distributivity of $\wedge$ over $\vee$ (A8) to push conjunctions inside disjunctions, and use (A11) and (A1) to turn negated literals into literals. $\square$

Lemma 9 (Standard disjunctive normal form)**.** *Every sentence is provably equivalent to one in standard disjunctive normal form.*

Proof. Given any sentence φ of $\mathcal{L}$, by the Normal Form Theorem (Theorem 8) φ is provably equivalent to one in disjunctive normal form. Using (A4–5), equivalence is preserved with the literals of each conjunct and the conjuncts of each disjunct placed according to the fixed order, and using associativity (A6–7) to associate brackets to the right. Equivalence is also preserved using (A3) to remove repeats of disjuncts, and since each conjunct is a conjunct of literals, we may use (A2) to remove repeats of literals. $\square$

2.7.2 The canonical model

Definition 10 (Canonical model)**.** *The canonical model $\mathcal{M}_c$ of IE is (X_c, ι_c) where $X_c = \mathbb{L}$ is the set of literals and $\iota_c : l \mapsto l$ the identity map.*

To prove completeness we need the following lemmata, showing that equivalence in the canonical model amounts to the syntactic notion of equiformity, or, symbol-for-symbol identity. We first need the following straightforward fact about pairwise union.

Fact 11. $A = \bigcup_{b \in A} \Cup_{c \in b} \{\{c\}\}$ *for any set A of nonempty sets.*

Proof. For each $b \in A$, let $b = \{c_i\}_{i \in I}$ for some index set I. As b is nonempty, $\Cup_{c \in b} \{\{c\}\} = \{\{c_0\} \cup \{c_1\} \cup \cdots : \{c_i\} \in \{\{c_i\}\}, i \in I\} = \{\{c_i : i \in I\}\} = \{b\}$. So $\bigcup_{b \in A} \Cup_{c \in b} \{\{c\}\} = \bigcup_{b \in A} \{b\} = A$. $\square$

Lemma 12. *For any φ in standard disjunctive normal form, $\varphi = \bigvee_{L \in i_c(\varphi)} \bigwedge_{l \in L} l$.*

Proof. $i_c(\varphi)$ is a set of nonempty sets, so by Fact 11, $i_c(\varphi) = \bigcup_{L \in i_c(\varphi)} \Cup_{l \in L} i_c(l)$, which is $\bigcup_{L \in i_c(\varphi)} \Cup_{l \in L} \{\{l\}\}$ by definition of i_c. Now note that $i_c(\varphi)$ is a finite set of finite sets of literals, again by construction of the canonical model. Then by construction of φ's standard disjunctive normal form, φ is $\bigvee_{L \in i_c(\varphi)} \bigwedge_{l \in L} l$. $\square$

Corollary 13. *For any sentences φ and ψ in standard disjunctive normal form, $i_c(\varphi) = i_c(\psi)$ just in case φ and ψ are equiform: $\varphi = \psi$.*

Proof. The right-to-left direction is trivial. For the left-to-right direction, note

$$\varphi = \bigvee_{L \in i_c(\varphi)} \bigwedge_{l \in L} l = \bigvee_{L' \in i_c(\psi)} \bigwedge_{l' \in L'} l' = \psi$$

with $i_c(\varphi) = i_c(\psi)$ giving the inner and Lemma 12 the outer identities. $\square$

The following lemma is then the key to the completeness proof, showing that theoremhood and truth in the canonical model coincide.

Lemma 14. *$\varphi \equiv \psi$ is a theorem just in case it is true in the canonical model $\mathcal{M}_c$.*

Proof. By soundness (Theorem 5), if $\varphi \equiv \psi$ is a theorem then it is true in the canonical model, since the canonical model is a model by Definition 2.

So suppose $\varphi \equiv \psi$ is true in the canonical model; that is, $i_c(\varphi) = i_c(\psi)$. By Lemma 9, φ is provably equivalent to a standard disjunctive normal form φ^*, and ψ is provably equivalent to a standard disjunctive normal form ψ^*. Then by soundness again, $\varphi \equiv \varphi^*$ and $\psi \equiv \psi^*$ are valid; in particular, they are true in the canonical model. Hence $i_c(\varphi^*) = i_c(\varphi) = i_c(\psi) = i_c(\varphi^*)$, so $\varphi^* = \psi^*$ by Corollary 13. Hence, by Fact 1, $\varphi^* \equiv \psi^*$ is a theorem. Then as φ^* is provably equivalent to φ and ψ^* to ψ, $\varphi \equiv \psi$ is a theorem by transitivity. $\square$

Theorem 15 (Completeness)**.** *Every validity of IE is a theorem.*

Proof. If $\varphi \equiv \psi$ is valid, it holds in $\mathcal{M}_c$, and so is a theorem by Lemma 14. $\square$

The proof of completeness also reveals a welcome result: IE is decideable.

Theorem 16 (Decidablility)**.** *Theoremhood of IE is decidable.*

Proof. As sentences are finite strings, it is efficient to determine the standard disjunctive normal forms of any sentences φ and ψ using the methods in the proofs of Theorems 8 and 9. It is furthermore efficient, of course, to determine whether their standard disjunctive normal forms are identical (in the syntactic sense of equiformity), and so by Corollary 13 whether they are equivalent in the canonical model, and hence by Lemma 14 whether $\varphi \equiv \psi$ is a theorem. $\square$

We have thus axiomatised a logic of imaginative content using the connective $\equiv$. The following section briefly considers a second connective one might use to articulate relationships between imaginative contents.

2.8 Imaginative inclusion

Imaginative equivalence is an interesting notion in its own right. One may, however, wish for a logic to describe a weaker notion of imaginative inclusion, whereby every way of imagining φ true is a way of imagining ψ true, but perhaps not vice versa. To that end, we may consider the language

$$\mathcal{L}_{\hookrightarrow} = \{\varphi \hookrightarrow \psi : \varphi, \psi \in \mathcal{L}\}$$

and give the truth condition for inclusions in $\mathcal{L}_{\hookrightarrow}$ like so.

Definition 17. *$\varphi \hookrightarrow \psi$ is true in a model (X, ι) of IE just in case $i(\varphi) \subseteq i(\psi)$.*

Nonetheless, imaginative inclusion turns out to be easily definable within IE already. We simply take $\varphi \hookrightarrow \psi$ to be $\varphi \vee \psi \equiv \psi$.

Fact 18 (Defining $\hookrightarrow$)**.** *$\varphi \vee \psi \equiv \psi$ is true in (X, ι) just in case $i(\varphi) \subseteq i(\psi)$.*

Proof. For any (X, ι), $i(\varphi \vee \psi) = i(\psi)$ iff $i(\varphi) \cup i(\psi) = i(\psi)$ iff $i(\varphi) \subseteq i(\psi)$. $\square$

By completeness, then, $\varphi \vee \psi \equiv \psi$ is a theorem of IE just in case $i(\varphi) \subseteq i(\psi)$ for every model (X, ι) of IE. And if we wished to add $\hookrightarrow$ explicitly to our language $\mathcal{L}_{\equiv}$, to preserve completeness we need only add the rule following rule (where double lines indicate interderivability).

$$\frac{\varphi \vee \psi \equiv \psi}{\varphi \hookrightarrow \psi}\ \text{Definition of} \hookrightarrow$$

Fact 19. *The following are derived rules of IE + (Definition of $\hookrightarrow$),*

$$\frac{\varphi \hookrightarrow \psi \qquad \psi \hookrightarrow \varphi}{\varphi \equiv \psi} \equiv \textit{introduction} \qquad \frac{\varphi \equiv \psi}{\varphi \hookrightarrow \psi} \equiv \textit{elimination}$$

$$\frac{\varphi \hookrightarrow \psi \qquad \psi \hookrightarrow \chi}{\varphi \hookrightarrow \chi}\ \textit{Transitivity of} \hookrightarrow \qquad \frac{\varphi \vee \psi \hookrightarrow \chi}{\varphi \hookrightarrow \chi} \hookrightarrow \textit{simplification}$$

and $\varphi \hookrightarrow \varphi \vee \psi$ is a theorem of IE + (Definition of $\hookrightarrow$).

Proof. Proven as Fact 60 of the appendix. $\square$

One might be tempted to go the other way, defining $\varphi \equiv \psi$ as $(\varphi \hookrightarrow \psi) \wedge (\psi \hookrightarrow \varphi)$. This definiens, however, is not a sentence of the language $\mathcal{L}_{\hookrightarrow}$ since the conjunction scopes over the imaginative connective $\hookrightarrow$. Indeed, the $\wedge$ appearing in the suggested definition seems to be a truth-conditional conjunction, unlike the unique 'imaginative' conjunction of IE. It is for this reason, in addition to ease of presentation, that we first presented a logic of imaginative equivalence rather than inclusion. But this is not a substantial issue. One can easily axiomatise a logic of imaginative inclusion by rewriting the axioms and rules of IE. (To be exact, we replace any axiom scheme $\varphi \equiv \psi$ with the two schema $\varphi \hookrightarrow \psi$ and $\psi \hookrightarrow \varphi$, replace any rule of the form $\varphi \equiv \psi/\chi$ with $\varphi \hookrightarrow \psi, \psi \hookrightarrow \varphi/\chi$ and $\varphi/\psi \equiv \chi$ with both $\varphi/\psi \hookrightarrow \chi$ and $\varphi/\chi \hookrightarrow \psi$.) To bring equivalence back into the picture, one adds the following rules.

$$\frac{\varphi \hookrightarrow \psi \qquad \psi \hookrightarrow \varphi}{\varphi \equiv \psi} \equiv \text{introduction} \qquad\qquad \frac{\varphi \equiv \psi}{\varphi \hookrightarrow \psi} \equiv \text{elimination}$$

In what follows, we will make use of both symbols $\equiv$ and $\hookrightarrow$, and for ease of reference will continue to refer to the logic as IE, even when we are talking about imaginative inclusion.

2.9 A logic of consistent imaginative equivalence

The logic IE of imaginative equivalence above does not give a special status to contradictions. For instance, a model of IE may well assign different contents to the sentences, 'It is raining and not raining' and, 'It is sunny and not sunny'. And while it is natural to think that, "we can occasionally have inconsistent conceptions" (Berto, 2017, 1282), one might also like a logic of imaginative content to assign inconsistent conceptions a special status.

One could even go so far as to say that sentences of the form $\varphi \wedge \neg\varphi$ are 'unimaginable,' so to speak, in the sense that their imaginative content is empty. This will show up at the level of imaginative equivalence since, for example, if $\varphi \wedge \neg\varphi$ and $\psi \wedge \neg\psi$ both have empty imaginative content, then they ought to be imaginatively equivalent.

To this end, we may formulate a logic of consistent imaginative equivalence, CIE. The task of this section is to formulate such a logic and prove its soundness and completeness.

2.9.1 Axiomatisation of CIE

The deductive system of CIE is given by the axiom schema and rules of inference of IE, in addition to the following axiom schema.

$$\varphi \wedge (\psi \wedge \neg\psi) \equiv \psi \wedge \neg\psi \tag{A14}$$

$$\varphi \vee (\psi \wedge \neg\psi) \equiv \varphi \tag{A15}$$

Let us say that a *contradiction* is any sentence of the form $\varphi \wedge \neg\varphi$. An immediate consequence of A14 is that all contradictions are equivalent according to CIE.

Fact 20. *$\varphi \wedge \neg\varphi \equiv \psi \wedge \neg\psi$ is a theorem of CIE, for any sentences φ, ψ of $\mathcal{L}$.*

Proof.

$$\dfrac{\dfrac{}{(\varphi \wedge \neg\varphi) \wedge (\psi \wedge \neg\psi) \equiv \varphi \wedge \neg\varphi}\,\text{A14}}{\varphi \wedge \neg\varphi \equiv (\varphi \wedge \neg\varphi) \wedge (\psi \wedge \neg\psi)}\,\text{Sym.} \qquad \dfrac{}{(\varphi \wedge \neg\varphi) \wedge (\psi \wedge \neg\psi) \equiv \psi \wedge \neg\psi}\,\text{A14}$$
$$\overline{\qquad\qquad\varphi \wedge \neg\varphi \equiv \psi \wedge \neg\psi\qquad\qquad}\,\text{Tran.}$$

□

Fact 20 licences one, if desired, to designate a sentence φ of $\mathcal{L}$ and metalinguistically define $\perp$ as an abbreviation of $\varphi \wedge \neg\varphi$ or to introduce a constant $\perp$ into the language of CIE, subject to the axiom $\perp \equiv \varphi \wedge \neg\varphi$ (by Fact 20 it does not matter which sentence we pick). We may then more conveniently express A14 and A15 like so.

$$\varphi \wedge \perp \equiv \perp \tag{A14$'$}$$

$$\varphi \vee \perp \equiv \varphi \tag{A15$'$}$$

2.9.2 Semantics of CIE

Onto the semantics, our models of CIE will be just those of IE.

Definition 21 (Model of CIE)**.** *A model of CIE is simply a model (X, ι) of IE.*

The semantic difference between IE and CIE instead comes in the function assigning imaginative contents to sentences. Given a set $C \subseteq X$ of imaginative contents, we call C contradictory just in case it contains both the imaginative content of an atomic sentence and its negation. The interpretation function of CIE is simply a restriction of IE's interpretation function to contents that are not contradictory. The idea is that a consistent logic of imaginative content should not assign any content to inconsistencies.

Definition 22 (Semantics of CIE)**.** *Given a model (X, ι) of CIE, the semantics of CIE results by extending the function ι to arbitrary sentences of $\mathcal{L}$ as follows.*

$$ci(\varphi) = \{C \in i(\varphi) : \textit{for no atomic sentence } p \textit{ are both } \iota(p), \iota(\neg p) \in C\}.$$

Analogous to IE, we say $\varphi \equiv \psi$ is true in a model (X, ι) just in case $ci(\varphi) = ci(\psi)$.

This is a very simple way to define CIE's semantics, one defined in terms of IE's semantics. As a result, CIE's semanics turns out to be non-compositional.[5] I am not too worried about the non-compositionality of the semantics for CIE presented above, for two reasons. The first is that one might intuitively take its

[5] As a counterexample to CIE's compositionality, consider $p \wedge \neg p$ and $\neg(p \wedge \neg p)$, and for simplicity, a model with $\iota(l) = l$ for every literal l. Then $i(p \wedge \neg p) = \{\{p, \neg p\}\}$, and so $ci(p \wedge \neg p) = \emptyset$. Similarly, $ci(q \wedge \neg q) = \emptyset$. Now, $i(\neg(p \wedge \neg p)) = \{\{p\}, \{\neg p\}, \{p, \neg p\}\}$, and so $ci(\neg(p \wedge \neg p)) = \{\{p\}, \{\neg p\}\}$. Then $ci(p \wedge \neg p) = ci(q \wedge \neg q)$, but $ci(\neg(p \wedge \neg p)) \neq ci(\neg(q \wedge \neg q))$.

semantics to be parasitic upon that of IE, holding that IE's semantics is closer to how we really interpret sentences in imaginative contexts, and that when we encounter a contradiction we invoke not purely semantic but pragmatic reasoning, such as a refusal to interpret the sentence based on a presupposition of consistency. The second reason is that, although the above semantics is very direct, one may use suitable tricks to force a compositional semantics. For as Pagin and Westerståhl put it, "it is always possible to enforce compositionality by *un*reasonable means" (Pagin and Westerståhl, 2010, 253). In Definition 22, I hope to have offered a non-compositional, but reasonable, semantics of CIE.

To prove soundness of CIE with respect to the above semantics, one first shows that every axiom of IE is valid by the semantics of CIE. To see this, note that for any axiom $\varphi \equiv \psi$ of IE, $i(\varphi) = i(\psi)$ holds by soundness of IE (Theorem 5), and so of course $ci(\varphi) = ci(\psi)$, the interpretation function ci being but a restriction of i. To prove the validity of the new axioms A14 and A15, one first proves the following straightforward lemma. Given a sentence φ of $\mathcal{L}$, say $i(\varphi) \subseteq \wp(X)$ is *contradictory* just in case every element $C \in i(\varphi)$ is contradictory, that is, $\iota(p), \iota(\neg p) \in C$ for some atomic sentence p.

Lemma 23. *Let (X, ι) be a model of IE (hence also of CIE). Given any sentence φ of $\mathcal{L}$, say $i(\varphi)$ is contradictory just in case every element $C \in i(\varphi)$ is contradictory, that is, $\iota(p), \iota(\neg p) \in C$ for some atomic sentence p. Then any sentences φ, ψ of $\mathcal{L}$,*

1. *$i(\varphi \wedge \neg\varphi)$ is contradictory.*
2. *If $i(\psi)$ is contradictory then so is $i(\varphi \wedge \psi)$.*

Proof. Proven in the appendix as Facts 63 and 61, respectively. □

To show the validity of (A14), first note that by the above lemma $i(\varphi \wedge (\psi \wedge \neg\psi))$ and $i(\psi \wedge \neg\psi)$ are both contradictory. Then as desired, we have,

$$ci(\varphi \wedge (\psi \wedge \neg\psi)) = \varnothing = ci(\psi \wedge \neg\psi).$$

As for (A15), note that since every element of $i(\psi \wedge \neg\psi)$ is contradictory,

$$ci(\varphi \vee (\psi \wedge \neg\psi)) = ci(\varphi) \cup ci(\psi \wedge \neg\psi) = ci(\varphi) \cup \varnothing = ci(\varphi).$$

To prove completeness, one follows much the same strategy as in IE. Let us say that a conjunctive normal form is *syntactically contradictory* just in case it contains p and $\neg p$ as conjuncts for some atomic sentence p. Then to prove completeness of CIE with respect to the semantics of Definition 22, we first use the axioms of IE to transform φ and ψ into their respective standard disjunctive normal forms φ^* and ψ^* (Theorem 9). Then, just as $ci(\varphi)$ eliminates the contradictory elements from $i(\varphi)$, we use A14 and A15 to eliminate the syntactically contradictory conjuncts from φ^* and ψ^*. Hence $ci_c(\varphi) = ci_c(\psi)$ holds just in case φ^* and ψ^*, with their contradictory conjuncts eliminated, are identical. Using the provable reflexivity of $\equiv$ (Fact 1), completeness follows at once.

Chapter 3

Imaginative propositional logic

So far we have only considered logics of imaginative equivalence, with no regard to the familiar truth-conditional meanings of negation, conjunction and disjunction. But this separation of imaginative and informative content is unwarranted. Firstly, we may want to make claims combining informative and imaginative content, as in (19).

(19) Aman is not a bachelor, but to imagine Aman as a bachelor is to imagine him unmarried. $\neg B \wedge (B \hookrightarrow \neg M)$

Nonetheless, from a logician's point of view, sentences dealing explicitly with the imagination are not the main motivation for an imaginative propositional logic. The impetus for a logic combining informative and imaginative content is its role as the launchpad for imaginative modal logic. And the ultimate reason for caring about imaginative modal logic is its ability to handle counterfactuals, which create modal contexts, in a way that understands counterfactual antecedents in terms of their imaginative content. For just as propositional logic is the base logic of propositional modal logic, we will need an imaginative propositional logic to build an imaginative modal logic (imaginative modal logic is the subject of chapter 4).

But we are getting ahead of ourselves. We first have to define an imaginative propositional logic, denoted IPL, and prove our main results of this chapter: the strong soundness and strong completeness of IPL. The following three sections (3.1–3.3) are devoted to describing IPL's language, syntax and semantics, respectively. This is all a straightforward blend of classical propositional logic and the logic of imaginative equivalence. The axiomatisation is concise and sound (Theorem 26) but not very intuitive, so in section 3.5 we present a more lucid—though more involved—rule called the 'ways of imagination' rule. To put it loosely, the rule that states that if two sentences are imaginatively equivalent, we may infer that every way of imagining one is equivalent to a way of the other. We then derive this rule from IPL's axiomatisation and use it to prove completeness (section 3.6) in the standard Henkin style.

3.1 Language of IPL

The language of IPL consists of two kinds of sentence. The reason for this distinction is that, as we saw, imaginative negation, conjunction and disjunction behave quite differently from their classical counterparts. For the sake of clarity, in this chapter we use $\alpha, \beta, \ldots$ for sentences in the language of IE, called the base language $\mathcal{L}$. Recall that the sentences of $\mathcal{L}$ are built up from atomic sentences like so.

$$\alpha ::= p \mid \neg\beta \mid \beta \wedge \gamma \mid \beta \vee \gamma$$

Then where α and β are sentences of $\mathcal{L}$, any sentence φ in the language of IPL, denoted $\mathcal{L}_{IPL}$, is built up using the following grammar.

$$\varphi ::= p \mid \neg\psi \mid \psi \wedge \chi \mid \psi \vee \chi \mid \alpha \equiv \beta$$

As before, we use $\alpha \hookrightarrow \beta$ to abbreviate $\alpha \vee \beta \equiv \beta$, and for any sentence of IPL not of the form $\alpha \equiv \beta$ we apply the usual abbreviations of propositional logic.[1]

3.2 Axiomatisation of IPL

The deductive system of IPL is given by combining those of classical propositional logic and of IE with three additional rules.

PL Axiom schema and rules of inference of classical propositional logic.

IE Axiom schema and rules of inference of IE.

Rules The following rules of inference, where $l, l_1, \ldots, l_n$ are literals and γ a conjunction of literals.

$$\frac{\alpha \hookrightarrow \gamma}{\gamma \hookrightarrow \alpha}\ \text{Literal symmetry} \qquad \frac{\gamma \hookrightarrow (\alpha_1 \vee \cdots \vee \alpha_n)}{(\gamma \hookrightarrow \alpha_1) \vee \cdots \vee (\gamma \hookrightarrow \alpha_n)}\ \vee\ \text{distribution}$$

$$\frac{(\alpha \wedge l) \hookrightarrow (l_1 \wedge \cdots \wedge l_n)}{(l \hookrightarrow l_1) \vee \cdots \vee (l \hookrightarrow l_n)}\ \wedge\ \text{distribution}$$

[1] Namely, we define $\varphi \to \psi := \neg\varphi \vee \psi$ and $\varphi \leftrightarrow \psi := (\varphi \to \psi) \wedge (\psi \to \varphi)$.

Some notes on the axiomatisation are in order. We make use of axiom *schema* of propositional logic, rather than axioms and a rule of uniform substitution, since uniform substitution is not an admissible rule of IPL. To see this, consider that while $p \equiv p \wedge p$ and $p \equiv \neg\neg p$ are theorems of IE, we do not want IPL to prove their substitution instances $\varphi \equiv \varphi \wedge \varphi$ and $\varphi \equiv \neg\neg\varphi$. These, as we have seen (Proposition 6), are not validities of IE. We therefore omit substitution and instead present the axiomatisation using axiom schema.

The rule of literal symmetry and $\vee$ distribution acknowledge the special status of literals, where each conjunction of literals represents a single way of imagining a sentence true. Literal symmetry says, intuitively, that if there is only one way of imagining γ true, and every way of imagining α true is a way of imagining γ true, then that one way must also be a way of imagining α true. And $\vee$ distribution states that if every way of imagining γ true is a way of imagining a disjunction true, then as there is only one way of imagining γ true, it must be a way of imagining one of the disjuncts true. Finally, $\wedge$ distribution says that if every way of $\alpha \wedge l$ is a way of imagining some literals $l_1 \wedge \cdots l_n$ true, then to imagine l must be to imagine at least one of those literals $l_1, \ldots, l_n$ true.

3.3 Semantics of IPL

The semantics of IPL is a simple pairing of the semantics of IE and classical propositional logic.

Definition 24 (Model of IPL)**.** *A model of IPL is a triple $w = (\mathcal{I}, X, \iota)$ where $\mathcal{I}$ is a model of propositional logic and (X, ι) a model of IE.*

The truth conditions for sentences of IPL are also a straightforward combination of those for propositional logic and IE. In the clauses below, as usual we let $w \models \varphi$ denote that φ is true in w.

Definition 25 (Semantics of IPL)**.** *Let $w = (\mathcal{I}, X, \iota)$ be a model of IPL and i the extension of ι to all sentences in $\mathcal{L}$ according to the semantics of IE (Definition 3).*

$$\begin{array}{lll}
w \models p & \textit{iff} & p \textit{ is true in } \mathcal{I}. \\
w \models \alpha \equiv \beta & \textit{iff} & i(\alpha) = i(\beta). \\
w \models \neg\psi & \textit{iff} & w \not\models \psi. \\
w \models \psi \wedge \chi & \textit{iff} & w \models \psi \textit{ and } w \models \chi. \\
w \models \psi \vee \chi & \textit{iff} & w \models \psi \textit{ or } \mathcal{M} \models \chi.
\end{array}$$

3.4 Soundness of IPL

Since classical propositional logic and IE are each sound with respect to their own semantics, to show soundness of IPL we need only show that the three new rules are truth-preserving in every model of IPL.

Theorem 26 (Soundness of IPL)**.** *Every theorem of IPL is true in every model of IPL.*

Proof. We prove the result in detail in the appendix (Theorem 64). The proof relies on fact that a conjunction of literals has only one imaginative content. □

While the three rules above are sound, they are not that intuitive. In the following section we present a single rule of inference that is harder to state, but I think much more intuitive. I call it the 'ways of imagination rule', for reasons that will soon become clear.

3.5 The ways of imagination rule

The semantics of IE interprets each sentence as a set of sets. For example, the sentence $p \vee (\neg q \wedge r)$ receives the semantic interpretation $\{\{\iota(p)\}, \{\iota(\neg q), \iota(r)\}\}$. Notice that each semantic object has two layers of structure. Given a sentence α of $\mathcal{L}$, the elements of $i(\alpha)$ correspond to the 'ways' of imagining α true. And for each way $A \in i(\alpha)$, the elements of A correspond to the 'imaginative contents' involved in that way of imagination.

This two-layered semantic picture also admits a syntactic representation. To show this, recall the methods used in the proof of IE's completeness. We put α in its standard disjunctive normal form α^*, which, like the semantic representation, also has two layers of structure. Each disjunct of α^* corresponds to a 'way' of imagining α true. And as disjunct is a conjunctive normal form (a conjunction of literals), each literal corresponds to an 'imaginative content' involved in a given way of imagining α true.

Now, IE's models operate over equivalences of literals, but IE's deductive system operates over equivalences of arbitrarily complex sentences. Our strategy for proving completeness will be to bridge that divide. While disjunctive normal forms allow us to reveal the semantic structure of a given sentence in the canonical model, they do not allow us to compare two sentences under some assumptions in models other than the canonical model. It is for this reason that we only proved the weak completeness of IE above (Theorem 15). Thankfully, however, the language of IPL has the expressive resources to give semantic equivalences of arbitrary models a syntactic form. To see this, note that we can decompose the identity $i(\alpha) = i(\beta)$ into the following two simpler kinds of equivalence, for any model (X, ι) of IE.

- Two ways of imagining $A, B \subseteq X$ are identical just in case every basic part of A is a basic part of B, and vice versa.
- Two sets of ways of imagining $i(\alpha), i(\beta) \subseteq \wp(X)$ are identical just in case every way of imagining in $i(\alpha)$ is identical to a way of imagining in $i(\beta)$, and vice versa.

Now, imaginative propositional logic does not represent set-theoretic identity. But then again it does not need to. Note that for any sets A and B, we may express $A \subseteq B$ in a more cumbersome way by saying that for every $a \in A$ there is a $b \in B$ such that $a = b$. The following notions then express the correspondences between, on the one hand, 'basic parts' of ways of imagining and conjunctive normal forms, and on the other, 'ways of imagining' and disjunctive normal forms.

- Two conjunctive normal forms are 'literally equivalent' just in case every literal of one is equivalent to a literal of the other, and vice versa.
- Two disjunctive normal forms are 'disjunctively equivalent' just in case every disjunct of one is literally equivalent to a disjunct of the other, and vice versa.

Let us now give precise definitions of the these notions of 'literal' and 'disjunctive' equivalence. For any conjunctive normal form $\gamma = l_1 \wedge \cdots \wedge l_m$, let $\textsc{lit}(\gamma) = \{l_1, \ldots, l_m\}$ denote the set of literals appearing in γ. For any conjunctive normal forms $\gamma, \delta \in \mathcal{L}$, we make use of the following abbreviations.

$$\gamma \hookrightarrow_{\textsc{lit}} \delta \;:=\; \bigwedge_{l \in \textsc{lit}(\gamma)} \bigvee_{l' \in \textsc{lit}(\delta)} (l \equiv l')$$

$$\gamma \equiv_{\textsc{lit}} \delta \;:=\; (\gamma \hookrightarrow_{\textsc{lit}} \delta) \wedge (\delta \hookrightarrow_{\textsc{lit}} \gamma)$$

Speaking intuitively, $\gamma \hookrightarrow_{\textsc{lit}} \delta$ states that every basic part of imagining γ true is equivalent to a basic part of imagining δ true. And $\gamma \equiv_{\textsc{lit}} \delta$ states that γ and δ are 'literal-for-literal equivalent', in the sense that every basic part of one is equivalent to a basic part of the other.

The literals appearing in a conjunctive normal form γ correspond to the 'basic parts' involved in imagining γ true. Moving up one syntactic level, the disjuncts appearing in a disjunctive normal form α correspond to the different 'ways' of imagining α true. For any disjunctive normal form (i.e. disjunction of conjunctive normal forms) $\alpha = \gamma_1 \vee \cdots \vee \gamma_n$, let $\textsc{dis}(\alpha) = \{\gamma_1, \ldots, \gamma_n\}$ denote the set of disjuncts appearing in α. Then for any disjunctive normal forms $\alpha, \beta \in \mathcal{L}$ we will also use the following abbreviations.

$$\alpha \hookrightarrow_{\textsc{dis}} \beta \;:=\; \bigwedge_{\gamma \in \textsc{dis}(\alpha)} \bigvee_{\delta \in \textsc{dis}(\beta)} (\gamma \equiv_{\textsc{lit}} \delta)$$

$$\alpha \equiv_{\textsc{dis}} \beta \;:=\; (\alpha \hookrightarrow_{\textsc{dis}} \beta) \wedge (\beta \hookrightarrow_{\textsc{dis}} \alpha)$$

The sentence $\alpha \hookrightarrow_{\textsc{dis}} \beta$ states, in effect, that every way of imagining α true is equivalent to a way of imagining β true.

Recall that for any sentence $\alpha \in \mathcal{L}$, α^* denotes the standard disjunctive normal form of α. Then let us finally use the following abbreviation.

$$\alpha \overset{*}{\equiv} \beta \;:=\; \alpha^* \equiv_{\textsc{dis}} \beta^*$$

With these abbreviations to hand we may succinctly formulate the 'ways of imagination rule' like so.

$$\frac{\alpha \equiv \beta}{\alpha \overset{*}{\equiv} \beta}\ \text{WoI}$$

Although more intuitive, the ways of imagination rule is harder to formulate than the three rules given in the axiomatisation of IPL above. This is why we formulated IPL in terms of the three shorter rules, instead of the initially more complicated—but eventually more succinct—was of imagination rule. Those three rules, however, are sufficient to derive the ways of imagination rule, as we show now.

3.5.1 Deriving the ways of imagination rule

We begin with the following lemma, which builds up incrementally to the ways of imagination rule.

Lemma 27. *The following are derived rules of IPL, for any conjunctive normal forms $\gamma, \delta \in \mathcal{L}$ and disjunctive normal forms $\alpha, \beta \in \mathcal{L}$.*

$$\frac{\gamma \hookrightarrow \delta}{\gamma \hookrightarrow_{\text{LIT}} \delta} \qquad \frac{\gamma \hookrightarrow \delta}{\gamma \equiv_{\text{LIT}} \delta} \qquad \frac{\alpha \hookrightarrow \beta}{\alpha \hookrightarrow_{\text{DIS}} \beta}$$

Proof. We work in the deductive system of IPL. Pick any conjunctive normal forms $\gamma, \delta \in \mathcal{L}$, i.e. γ and δ are of the form $\gamma = l_1 \wedge \cdots \wedge l_m$ and $\delta = l'_1 \wedge \cdots \wedge l'_n$.

Assume $\gamma \hookrightarrow \delta$ as a premise. Then apply $\wedge$ distribution.

$$\frac{(l_1 \wedge \cdots \wedge l_{m-1}) \wedge l_m \hookrightarrow (l'_1 \wedge \cdots \wedge l'_n)}{(l_m \hookrightarrow l'_1) \vee \cdots \vee (l_m \hookrightarrow l'_n)}\ \wedge \text{ distribution}$$

By associativity and communitativity of $\wedge$ (A5, A7), we may apply $\wedge$ distribution m-many times to derive $\bigvee_{l' \in \text{LIT}(\delta)}(l \hookrightarrow l')$ for each $l \in \text{LIT}(\gamma)$. Then from literal symmetry we derive $\bigvee_{l' \in \text{LIT}(\delta)}(l \equiv l')$ for each $l \in \text{LIT}(\gamma)$. So by propositional logic derive $\bigwedge_{l \in \text{LIT}(\gamma)} \bigvee_{l' \in \text{LIT}(\delta)}(l \equiv l')$. That is, $\gamma \hookrightarrow_{\text{LIT}} \delta$.

From $\gamma \hookrightarrow \delta$ by literal symmetry we derive $\delta \hookrightarrow \gamma$. Then by the above we derive $\gamma \hookrightarrow_{\text{LIT}} \delta$ and $\delta \hookrightarrow_{\text{LIT}} \gamma$, and so derive $\gamma \equiv_{\text{LIT}} \delta$ by propositional logic.

Now pick any disjunctive normal forms $\alpha, \beta \in \mathcal{L}$. Then α and β are of the form $\alpha = \gamma_1 \vee \cdots \vee \gamma_m$ and $\beta = \delta_1 \vee \cdots \vee \delta_n$. Assume $\alpha \hookrightarrow \beta$ as a premise.

Recall that $\hookrightarrow$-simplification is a derived rule of IE (Fact 19), so also of IPL.

$$\frac{\alpha \vee \alpha' \hookrightarrow \beta}{\alpha \hookrightarrow \beta}\ \hookrightarrow \text{simplification}$$

As α is a disjunctive normal form, it is of the form $\alpha = \gamma_1 \vee \cdots \vee \gamma_m$. We then apply $\vee$ elimination.

$$\frac{\gamma_1 \vee (\gamma_2 \vee \cdots \vee \gamma_m) \hookrightarrow \beta}{\gamma_1 \hookrightarrow \beta} \vee \text{ elimination}$$

Using associativity and communitativity of $\vee$ (A4, A6), we apply $\vee$ elimination m-many times to derive $\gamma \hookrightarrow \beta$ for each $\gamma \in \text{DIS}(\alpha)$. Pick any such γ. As β is also a disjunctive normal form, it is of the form $\beta = \delta_1 \vee \cdots \vee \delta_n$. And as γ is a conjunctive normal form, we may apply $\vee$ distribution.

$$\frac{\gamma \hookrightarrow (\delta_1 \vee \cdots \vee \delta_n)}{(\gamma \hookrightarrow \delta_1) \vee \cdots \vee (\gamma \hookrightarrow \delta_n)} \vee \text{ distribution}$$

Then by propositional logic we derive $\bigvee_{\delta \in \text{DIS}(\beta)} (\gamma \equiv_{\text{LIT}} \delta)$. As γ we arbitrary, by propositional logic again we derive $\bigwedge_{\gamma \in \text{DIS}(\alpha)} \bigvee_{\delta \in \text{DIS}(\beta)} \gamma \hookrightarrow_{\text{DIS}} \delta$.

□

From this lemma it is rather straightforward to derive the WoI rule in full.

Corollary 28. *The ways of imagination rule WoI is a derived rule of IPL.*

Proof. We work in the deductive system of IPL. Pick any $\alpha, \beta \in \mathcal{L}$ and assume $\alpha \equiv \beta$ as a premise. We have to derive $(\alpha^* \hookrightarrow_{\text{DIS}} \beta^*) \wedge (\beta^* \hookrightarrow_{\text{DIS}} \alpha^*)$.

By IE's normal form theorem, $\alpha \equiv \alpha^*$ and $\beta \equiv \beta^*$ are theorems of IE, so also of IPL. Then from $\alpha \equiv \beta$ by transitivity we derive $\alpha^* \equiv \beta^*$. Thus by Fact 19 we derive $\alpha^* \hookrightarrow \beta^*$ and $\beta^* \hookrightarrow \alpha^*$. Then by Proposition 27, we respectively derive $\alpha^* \hookrightarrow_{\text{DIS}} \beta^*$ and $\beta^* \hookrightarrow_{\text{DIS}} \alpha^*$, whence we have $\alpha \overset{*}{\equiv} \beta$ by propositional logic. □

As the three short rules (literal symmetry, $\vee$ and $\wedge$ distribution) are truth-preserving with respect to IPL's semantics (Theorem 26), the above corollary tells us that the ways of imagination rule is also truth-preserving with respect to IPL's semantics. This result means, in effect, that we have two axiomatisations of IPL at our disposal. We can add the three short rules used above, or we can add the single ways of imagination rule directly.

It remains to show that our axiomatisation is complete. We prove this using the ways of imagination rule since it furnishes a particularly elegant proof in the standard Henkin style.

3.6 Completeness of IPL

In this section we prove that IPL is complete with respect to the semantics presented above. Our proof strategy is to extend Leon Henkin's canonical model construction for propositional logic.

First designate a sentence φ of $\mathcal{L}_{IPL}$ and define $\perp$ as $\varphi \wedge \neg\varphi$. This conjunction, appearing outside the scope of the imaginative connective $\equiv$, is classical (hence not paraconsistent), so any φ at all will do. For any set Γ of sentences of IPL, we call Γ *consistent* just in case $\perp$ is not derivable from the members of Γ using the axiom schema and rules of IPL's deductive system. To begin, we need the usual definition of a maximally consistent set of sentences.

Definition 29 (IPL-maximal consistency). *A set Φ of sentences of $\mathcal{L}_{IPL}$ is an IPL-maximally consistent set (for short: an IPL-*MCS*) just in case (i) Φ is consistent but (ii) for every sentence $\varphi \notin \Phi$ of $\mathcal{L}_{IPL}$, $\Phi \cup \{\varphi\}$ is inconsistent.*

A standard adaptation of Lindenbaum's lemma shows that every consistent set Γ of sentences of $\mathcal{L}_{IPL}$ can be extended to an IPL-MCS $\Phi \supseteq \Gamma$.

Lemma 30 (Truth lemma). *For any IPL-*MCS *Φ, $\varphi, \psi \in \mathcal{L}_{IPL}$ and $\alpha, \beta \in \mathcal{L}$,*

($\neg$) *$\neg\varphi \in \Phi$ iff $\varphi \notin \Phi$.*

($\wedge$) *$\varphi \wedge \psi \in \Phi$ iff $\varphi \in \Phi$ and $\psi \in \Phi$.*

($\vee$) *$\varphi \vee \psi \in \Phi$ iff $\varphi \in \Phi$ or $\psi \in \Phi$.*

($\equiv$) *$\alpha \equiv \beta \in \Phi$ iff $\alpha \overset{*}{\equiv} \beta \in \Phi$.*

Proof. Since IPL's deductive system subsumes that of classical propositional logic, clauses ($\neg$), ($\wedge$) and ($\vee$) follow from a standard argument from maximal consistency of Φ (e.g. Van Dalen, 1994, §1.5). The left-to-right direction of ($\equiv$) follows from the ways of imagination rule (WoI). For if we had $\alpha \equiv \beta \in \Phi$ but $\alpha \overset{*}{\equiv} \beta \notin \Phi$ by maximality of Φ, $\Phi \cup \{\alpha \overset{*}{\equiv} \beta\}$ would be inconsistent, in which case Φ would itself be inconsistent by WoI, contradicting our assumption.

For the right-to-left direction, suppose $\alpha \overset{*}{\equiv} \beta \in \Phi$. By the disjunctive normal form theorem for IE, $\alpha \equiv \alpha^*$ and $\beta \equiv \beta^*$ are theorems of IE, and so also of IPL. Then $\alpha \equiv \alpha^* \in \Phi$ and $\beta \equiv \beta^* \in \Phi$. By transitivity, then, to show $\alpha \equiv \beta \in \Phi$ it suffices to show that $\alpha^* \equiv \beta^* \in \Phi$.

Consider any disjunct γ of α^*. Since $\alpha \overset{*}{\equiv} \beta \in \Phi$, by ($\wedge$) and ($\vee$) we have both $(\gamma \hookrightarrow_{\text{LIT}} \delta), (\delta \hookrightarrow_{\text{LIT}} \gamma) \in \Phi$ for some disjunct δ of β^*. Now consider any literal l of γ. Then by ($\wedge$) and ($\vee$) again, $(l \equiv l') \in \Phi$ for some literal l' of δ.

Now, when we constructed each sentence's standard conjunctive normal form we fixed a linear ordering $\leq$ over the literals. So consider the $\leq$-least literal l_0 such that $l \equiv l_0 \in \Phi$. As γ is a conjunctive normal form, l does not occur under the scope of $\neg$ in γ, so by substitution (Proposition 7), $\gamma \equiv \gamma[{}^{l_0}/_{l}] \in \Phi$. Continue applying substitution to replace every literal of γ with the $\leq$-least literal l_k such that $l \equiv l_k \in \Phi$. (Since $l \equiv l \in \Phi$, such a literal is guaranteed to exist.) Call the result γ_δ. Then every literal of γ_δ is a literal of δ and $\gamma_\delta \equiv \delta \in \Phi$.

Similarly, use substitution and the fact that $\delta \hookrightarrow_{\textsc{lit}} \gamma \in \Phi$ to construct δ_γ with $\delta_\gamma \equiv \gamma \in \Phi$. Then by construction of γ_δ and δ_γ, they are composed of the exact same literals, i.e. $\gamma_\delta = \delta_\gamma$, and so by reflexivity of $\equiv$, $\gamma_\delta \equiv \delta_\gamma$ is a theorem of IE. Hence $(\gamma \equiv \delta_\gamma), (\delta_\gamma \equiv \gamma_\delta), (\gamma_\delta \equiv \delta) \in \Phi$, whence by transitivity $(\gamma \equiv \delta) \in \Phi$.

Similarly, in place of substituting literals one may substitute whole conjunctive normal forms in α^* for conjunctive normal forms in β^*. When we constructed the standard disjunctive normal forms we fixed an ordering over the standard conjunctive normal forms. So we may construct α_β and β_α with $\alpha_\beta = \beta_\alpha$. Similar to the above, therefore, $\alpha^* \equiv \beta^* \in \Phi$, as required. $\square$

Definition 31 (Canonical model for Φ)**.** *Fix a linear order $\leq$ of literals and let Φ be an IPL-*mcs*. The canonical model for Φ is defined to be $w_\Phi = (\mathcal{I}_\Phi, X_\Phi, \iota_\Phi)$, where*

1. *$\mathcal{I}_\Phi$ makes p true iff $p \in \Phi$.*
2. *X_Φ is the set of literals.*
3. *For any literal l, $\iota_\Phi(l)$ is the least literal l' w.r.t. $\leq$ such that $l \equiv l' \in \Phi$.*

This definition ensures that $l \equiv l' \in \Phi$ iff $i_\Phi(l) = i_\Phi(l')$ for any literals l, l'. Our next lemma shows that membership in an IPL-mcs Φ amounts to truth in the canonical model for Φ. We build up to the result incrementally.

Lemma 32. *Let Φ be any IPL-*mcs*, γ, $\delta \in \mathcal{L}$ any conjunctive normal forms and α, $\beta \in \mathcal{L}$ any disjunctive normal forms. Then we have the following equivalences.*

1. *$\gamma \hookrightarrow_{\textsc{lit}} \delta \in \Phi$ iff for every $L \in i_\Phi(\gamma)$ there is an $L' \in i_\Phi(\delta)$ with $L \subseteq L'$.*
2. *$\gamma \equiv_{\textsc{lit}} \delta \in \Phi$ iff $i_\Phi(\gamma) = i_\Phi(\delta)$.*
3. *$\alpha \hookrightarrow_{\textsc{dis}} \beta \in \Phi$ iff $i_\Phi(\alpha) \subseteq i_\Phi(\beta)$.*
4. *$\alpha \equiv_{\textsc{dis}} \beta \in \Phi$ iff $i_\Phi(\alpha) = i_\Phi(\beta)$.*

Hence for any $\alpha, \beta \in \mathcal{L}$ whatsoever, $\alpha \overset{*}{\equiv} \beta \in \Phi$ iff $i_\Phi(\alpha) = i_\Phi(\beta)$.

Proof. Let γ, $\delta \in \mathcal{L}$ any conjunctive normal forms and α, $\beta \in \mathcal{L}$ any disjunctive normal forms. (1) By IE's semantic clause for conjunction, for some set of literals L, L' we have $i_\Phi(\gamma) = \{L\} = \{\{\iota_\Phi(l) : l \in \textsc{lit}(\gamma)\}\}$ and $i_\Phi(\delta) = \{L'\} = \{\iota_\Phi(l') : l' \in \textsc{lit}(\delta)\}$. Thus for (1) it suffices to show that $\gamma \hookrightarrow_{\textsc{lit}} \delta \in \Phi$ iff $L \subseteq L'$.

Arguing left-to-right, suppose $\gamma \hookrightarrow_{\textsc{lit}} \delta \in \Phi$ and pick any $l \in L$. To show that $l \in L'$, we have to show that $l = \iota_\Phi(l')$ for some $l' \in \textsc{lit}(\delta)$. As $l \in L$, $l = \iota_\Phi(l_0)$ for some $l_0 \in \textsc{lit}(\gamma)$. Then as $\gamma \hookrightarrow_{\textsc{lit}} \delta \in \Phi$, by maximal consistency of Φ (clauses $(\wedge)$ and $(\vee)$ of Lemma 30), $l_0 \equiv l' \in \Phi$ for some literal $l' \in \textsc{lit}(\delta)$. Hence, by definition of ι_Φ and linearity $\leq$ (the fixed order over the literals), $\iota_\Phi(l_0) = \iota_\Phi(l')$. And since $l = \iota_\Phi(l_0)$, we have $l = \iota_\Phi(l')$.

Conversely, suppose $L \subseteq L'$. By IE's semantic clause for conjunction, for some sets of literals L, L' we have $i_\Phi(\gamma) = \{L\} = \{\{\iota_\Phi(l) : l \in \textsc{lit}(\gamma)\}\}$ and $i_\Phi(\delta) = \{L'\} = \{\iota_\Phi(l') : l' \in \textsc{lit}(\delta)\}$. Pick any $l \in \textsc{lit}(\gamma)$. To show that $\gamma \hookrightarrow_{\textsc{lit}} \delta$ we have to show that $l \equiv l' \in \Phi$ for some $l' \in \textsc{lit}(\delta)$. As $l \in \textsc{lit}(\gamma)$, and $L \subseteq L'$, $\iota_\Phi(l) \in L'$. So $\iota_\Phi(l) = \iota_\Phi(l')$ for some literal l' of δ. Then by reflexivity, $\iota_\Phi(l) \equiv \iota_\Phi(l') \in \Phi$. Now, by construction of $\iota_\Phi(l)$ and $\iota_\Phi(l')$, $l \equiv \iota_\Phi(l) \in \Phi$ and $\iota_\Phi(l') \equiv l' \in \Phi$, so by transitivity $l \equiv l' \in \Phi$, as required.

(2) An easy consequence of (1): $\gamma \hookrightarrow_{\textsc{lit}} \delta \in \Phi$ iff $L \subseteq L'$, for $i_\Phi(\gamma) = \{L\}$ and $i_\Phi(\delta) = \{L'\}$. So $\gamma \equiv_{\textsc{lit}} \delta \in \Phi$ iff $L = L'$ iff $\{L\} = \{L'\}$ iff $i_\Phi(\gamma) = i_\Phi(\delta)$.

(3) Note that $i_\Phi(\alpha) = \{A_\gamma : \gamma \in \textsc{dis}(\alpha)\}$ where we define $\{A_\gamma\} := i_\Phi(\gamma)$ for every $\gamma \in \textsc{dis}(\alpha)$. Likewise, $i_\Phi(\beta) := \{B_\delta : \delta \in \textsc{dis}(\beta)\}$ where $\{B_\delta\} = i_\Phi(\delta)$ for every $\delta \in \textsc{dis}(\beta)$. Now suppose $\alpha \hookrightarrow_{\textsc{dis}} \beta \in \Phi$ and pick any $A \in i_\Phi(\alpha)$. We have to show that $A \in i_\Phi(\beta)$, that is, $A = B$ for some $B \in i_\Phi(\beta)$. As $A \in i_\Phi(\alpha)$, $A = A_\gamma$ for some $\gamma \in \textsc{dis}(\alpha)$. Then as $\alpha \hookrightarrow_{\textsc{dis}} \beta \in \Phi$, $\gamma \equiv_{\textsc{lit}} \delta$ for some $\delta \in \textsc{dis}(\beta)$. Then by (2), $i_\Phi(\gamma) = i_\Phi(\delta)$. As $\delta \in \textsc{dis}(\beta)$, $i_\Phi(\delta) = \{B\}$ for some $B \in i_\Phi(\beta)$. Thus, $\{A_\gamma\} = i_\Phi(\gamma) = i_\Phi(\delta) = \{B\}$, so $A = B$, as required.

Conversely, suppose $i_\Phi(\alpha) \subseteq i_\Phi(\beta)$ and pick any $\gamma \in \textsc{dis}(\alpha)$. We have to find some $\delta \in \textsc{dis}(\beta)$ with $\gamma \equiv_{\textsc{lit}} \delta \in \Phi$. Since $\gamma \in \textsc{dis}(\alpha)$, $i_\Phi(\gamma) = \{A_\gamma\}$ with $A_\gamma \in i_\Phi(\alpha)$. And as $i_\Phi(\alpha) \subseteq i_\Phi(\beta)$, $A_\gamma \in i_\Phi(\beta)$. By construction of $i_\Phi(\beta)$, $A_\gamma = B_\delta$ for some $\delta \in \textsc{dis}(\beta)$. Thus $i_\Phi(\gamma) = \{A_\gamma\} = \{B_\delta\} = i_\Phi(\delta)$, so by (2) $\gamma \equiv_{\textsc{lit}} \delta \in \Phi$ and we're done.

(4) $\alpha \equiv_{\textsc{dis}} \beta \in \Phi$ iff $\alpha \hookrightarrow_{\textsc{dis}} \beta, \beta \hookrightarrow_{\textsc{dis}} \alpha \in \Phi$, so (4) is immediate from (3).

Now by IE's standard normal form theorem (Theorem 9), IE proves $\alpha \equiv \alpha^*$, so $i_\Phi(\alpha) = i_\Phi(\alpha^*)$. Similarly, $i_\Phi(\beta) = i_\Phi(\beta^*)$. Then as $\alpha \overset{*}{\equiv} \beta \in \Phi$ iff $\alpha^* \equiv_{\textsc{dis}} \beta^* \in \Phi$, by (4) this holds iff $i_\Phi(\alpha^*) = i_\Phi(\beta^*)$, which by the above holds iff $i_\Phi(\alpha) = i_\Phi(\beta)$. □

Theorem 33. *For every IPL-*MCS *Φ there is a model $w = (\mathcal{I}, X, \iota)$ of IPL where, for any sentence φ of IPL, φ is true in $\mathcal{M}$ iff $\varphi \in \Phi$.*

Proof. Given any IPL-MCS Φ, consider its canonical model w_Φ (Definition 31) and pick any sentence φ of $\mathcal{L}_{IPL}$. We show by induction on the complexity of φ that φ is true in w_Φ iff $\varphi \in \Phi$. A standard argument from Lemma 30, respectively clauses $(\neg)$, $(\wedge)$ and $(\vee)$, shows the claim for φ of the form $\neg\psi$, $\psi \wedge \chi$ and $\psi \vee \chi$.

So suppose φ is $\alpha \equiv \beta$. Then we have the following chain of equivalences.

$$\begin{aligned} \alpha \equiv \beta \in \Phi \text{ iff } & \alpha \overset{*}{\equiv} \beta \in \Phi && \text{(Truth lemma 30, } \equiv) \\ \text{iff } & i_\Phi(\alpha) = i_\Phi(\beta) && \text{(Lemma 32)} \\ \text{iff } & \alpha \equiv \beta \text{ is true in } w_\Phi && \text{(Semantic definition of } \equiv) \end{aligned}$$

□

Theorem 34 (Strong completeness of IPL)**.** *Let φ be a sentence and Γ a set of sentences of $\mathcal{L}_{IPL}$. If $\Gamma \models \varphi$ then $\Gamma \vdash \varphi$.*

Proof. Contrapositively, suppose $\Gamma \nvdash \varphi$. Then $\Gamma \cup \{\neg\varphi\}$ is consistent, so by Lindenbaum's lemma there is an IPL-MCS Φ with $\Gamma \cup \{\neg\varphi\} \subseteq \Phi$. By Theorem 33, there is a model w_Φ of IPL making every member of Φ true, so by IPL's semantic clause for negation, w_Φ does not make φ true. Therefore, there is a model of IPL making every member of Γ true but not φ; that is to say, $\Gamma \not\models \varphi$. $\square$

One advantage to proving completeness with the ways of imagination rule is that the proof may easily be adapted to equivalential logics other than IE. For instance, one may define a consistent imaginative propositional logic CIPL based on the logic of consistent imaginative equivalence CIE. What changes in the proof is simply the standard disjunctive normal form: one defines a 'ways of consistently imagining rule' rule WoCI, with CIE's standard disjunctive normal form used in place of IE's. One may then show completeness for CIPL using the exact same techniques as above. Since, however, that would involve retreading covered ground, we will not give the proof here.

Now that we have a strongly axiomatised a propositional logic blending informative and imaginative content, we can use this logic as the basis for imaginative propositional modal logics that interpret antecedents in terms of their imaginative content. This is the task of our next chapter.

Chapter 4

Imaginative modal logic

We are all familiar with the idea that to evaluate counterfactuals, in general, we have to look beyond the actual world. While there are assertable counterfactuals with a true antecedent—the classic example is Anderson's, "If Jones had taken arsenic, he would have shown just exactly those symptoms which he does in fact show" (Anderson, 1951)—the point stands that counterfactuals create modal contexts. Evaluating counterfactuals, we do not just think about the material facts of the actual world. Thinking counterfactually, we imagine things otherwise.

Since counterfactuals create modal contexts, a logician would readily expect the techniques of modal logic to have something to contribute to the study of counterfactuals. This is the contribution of the present chapter. We set up the framework for a number of 'imaginative modal logics', based on imaginative propositional logic, which interpret sentences in terms of their informative and imaginative content (section 4.1). We then consider some intuitively plausible inference patterns of counterfactual logic (4.2) and formalise each within imaginative modal logic (4.3). We stop briefly to consider one rule of inference that some have proposed and give a counterexample to its intuitive validity (4.4). Then it is on to our main technical result of this chapter: the strong soundness and completeness of a variety of imaginative modal logics (4.5). We end with one aspect of our counterfactual interpretation missing from the model theory of imaginative modal logic: the passage of time (4.6). And so to work.

4.1 Language and semantics of IML

The language of IML draws on an idea going back to Chellas (1975), where we introduce one modal operator for each sentence of the language.[1]

[1]Such a modal language for conditionals is also used by Priest (2008, §5.3) and Berto (2017, 2018).

Definition 35 (Language of imaginative modal logic, $\mathcal{L}_{IML}$)**.** *φ is a sentence in the language of IML, written $\varphi \in \mathcal{L}_{IML}$, just in case it is of one of the following forms.*

$$\varphi ::= p \mid \neg\psi \mid \psi \vee \chi \mid \psi \wedge \chi \mid \alpha \equiv \beta \mid [\varphi]\psi \mid \langle\varphi\rangle\psi$$

where α and β are sentences of $\mathcal{L}$, and ψ and χ sentences of $\mathcal{L}_{IML}$.

We then introduce the familiar idea of a normal modal logic. The normal modal logics are those interpretable over relational structures.

Definition 36 (Normal imaginative modal logic)**.** *Let Λ be a set of sentences of imaginative modal logic. We say Λ is normal iff for every $\varphi, \psi, \chi \in \mathcal{L}_{IML}$,*

K schema	*Λ contains $[\varphi](\psi \to \chi) \to ([\varphi]\psi \to [\varphi]\chi)$.*
Dual schema	*Λ contains $\langle\varphi\rangle\psi \leftrightarrow \neg[\varphi]\neg\psi$.*
Necessitation	*If Λ contains ψ then it also contains $[\varphi]\psi$.*
Modus ponens	*If Λ contains φ and $\varphi \to \psi$ then it also contains ψ.*
IPL	*Λ is closed under IPL's axiom schema and rules of inference.*

We say a sentence φ of $\mathcal{L}_{IML}$ is a theorem of Λ just in case it is a member of Λ. Let IK *denote the smallest normal imaginative modal logic (*IK *for "Imaginative Kripke").*

The astute reader will recognise that one traditional rule is missing from our definition of normal modal logics: uniform substitution (the rule that if Λ contains φ then it also contains every sentence that results by uniformly replacing every atomic sentence in φ with a sentence of IML). We use axiom schema in place of uniform substitution since, as we saw in the axiomatisation of IPL (3.2), substitution is not an admissible rule of IPL, and a fortiori neither of IML.

Let us now present the semantics of IML. It will prove handy to have the semantics in place when we later introduce axioms for IML.

Definition 37 (Model of IML)**.** *A model of IML is a triple $(W, \{R_\varphi\}_{\varphi \in \mathcal{L}_{IML}}, V)$ where*

1. *W is a nonempty set.*
2. *$\{R_\varphi\}_{\varphi \in \mathcal{L}_{IML}}$ is a set of binary relations $R_\varphi \subseteq W \times W$, one for each $\varphi \in \mathcal{L}_{IML}$.*
3. *$V : w \mapsto (\mathcal{I}, X, \iota)$ is a valuation assigning to each point $w \in W$ a model of IPL.*

For any $w \in W$ with $V(w) = (\mathcal{I}, X, \iota)$, we will often simply write $w = (\mathcal{I}, X, \iota)$.

The intended interpretation of each accessibility relation R_φ is that v is R_φ-accessible from w just in case v is a way of imagining φ true at w. This point is rather cryptic at the moment, but will become clearer when we consider in detail the axiomatisation of various imaginative modal logics (4.3).

The above definitions strive for generality by providing one relation R_φ for each sentence φ of the entire language of imaginative modal logic: there are no restrictions on what sentences may feature as antecedents. Thus, for instance, even sentences about imaginative content and counterfactuals themselves get their own accessibility relation. This allows such sentences to be embedded as antecedents within more complex counterfactuals, as examples (20)–(22) illustrate.

(20) ALI: But just imagine if I were rich. I would have a yacht, champagne with gold flakes for breakfast...

VIV: If to imagine you rich were to imagine you happy, I would tell you to get rich. But it doesn't, so I won't.

$$[\text{Ali rich} \hookrightarrow \text{Ali happy}]\text{Viv tells Ali to get rich}$$

(21) TIM: If imagining time flowing were the just same as imagining entropy increasing, then a violation of the second law of thermodynamics would be inconceivable.

$$[\text{Time flows} \equiv \text{Entropy increases}]\text{Violation inconceivable}$$

(22) ZARA: Ok, so you didn't know getting a puppy is this much work. But even if you had known, you're so impulsive I bet you would've got one anyway. Now, if you were sensible—by which I mean, if you had known how much care a puppy needs you wouldn't have got one—and also, if you had bothered to find out how much work it takes, then I wouldn't be left cleaning up Toby's mess!

$$[[\text{Know}]\neg\text{Get puppy} \wedge \text{Know}]\neg\text{Get puppy}$$

Turning now to IML's truth conditions, we add the following clause to the semantics of IPL, which is standard for any box modality.

Definition 38 (Semantics of IML)**.** *Let $\mathcal{M} = (W, \{R_\varphi\}_{\varphi \in \mathcal{L}_{IML}}, V)$ be a model of IML and $\varphi, \psi \in \mathcal{L}_{IML}$. The semantics of IML adds to that of IPL the following clause.*

$$\mathcal{M}, w \models [\varphi]\psi \quad \textit{iff} \quad \textit{for every } v \in W, \textit{ if } wR_\varphi v \textit{ then } \mathcal{M}, v \models \psi.$$

Let $[\![\varphi]\!]_\mathcal{M} = \{w \in W : \mathcal{M}, w \models \varphi\}$ be the set of worlds where φ is true in $\mathcal{M}$.

This clause captures the thought that a counterfactual is true just in case the consequent holds under every way of imagining the antecedent true.[2] Of course, we have yet to consider what rules govern such acts of imagination. So let us now turn to some intuitive inference patterns that a logic of counterfactuals ought to validate.

[2]For any reader worried that this talk of imagination sounds overly subjective for a semantics of counterfactuals, see section 5.1.

4.2 Some inference patterns of counterfactual logic

This section is devoted to considering some examples of intuitive patterns of counterfactual inference. Some of these inferences have familiar names, though where they do not we provide a label to refer back to them later.

Simplification of disjunctive antecedents (SDA) (so-called by Nute, 1975)

(23) If either Mary or the Jones twins had tutored Johnny, he would have passed his algebra course.

a. So if Mary had tutored Johnny, he would have passed.

Creary and Hill (1975), see also Fine (1975)

SDA has many friends. It is explicitly endorsed, for example, by Nute (1975) Ellis et al. (1977), Fine (2012a), Starr (2014) and Willer (2018), who offer to explain its validity in the face of well-known counterexamples, such as the classic (24) from McKay and van Inwagen (1977).

(24) If Spain had fought with the Allies or the Axis, they would have fought with the Axis.

a. # So, if Spain had fought with the Allies, they would have fought with the Axis.

A popular defence of SDA's validity is that the rule is pragmatically blocked when one of the antecedent's disjuncts is not a relevant alternative (Willer, 2015). But Lassiter also notices that this defence does not extend to counterfactuals whose consequent features complex sentential operators such as *probably*, *usually* and *normally*, as in (25).

(25) If a classical musician switched to playing jazz or hip-hop she would normally—but not necessarily—switch to playing jazz.

a. # So, if a classical musician switched to playing hip-hop she would normally—but not necessarily—switch to playing jazz.

(Lassiter 2018, item 12)

Evidently, a speaker of (25) takes switching to hip-hop as a relevant alternative to switching to jazz, but the unacceptable inference to (25a) shows that SDA should not apply here.

Lassiter concludes that counterfactuals with complex sentential operators pose a great challenge to theories validating SDA *tout court*. An alternative proposal is that we have two ways of reasoning with multiple counterfactual alternatives: one is to hypothetically assume each alternative in turn and check for the consequent; the other, to weigh up the plausibility of the alternatives against one another. The former interpretation individuates counterfactual alternatives and validates SDA, while the latter compares counterfactual alternatives and does not validate SDA. If this 'two-interpretations' proposal is to work, one must identify the features of counterfactuals that explain why each receives the interpretation that it does; for example, why (23) receives the individuating interpretation but (24) the comparative interpretation.

An adequate defence of the two-interpretations proposal is a delicate task, one taking us too far from the main trajectory of this essay. Instead we will simply put up our hands in restricting the remarks below to counterfactuals that are interpreted by individuating each counterfactual alternative in turn.

Simplification of negated conjunctive antecedents (SNCA)

(26) If Nixon and Agnew had not both resigned, Ford would not have been president.

a. So if Nixon had not resigned, Ford would not have been president.
b. So if Agnew had not resigned, Ford would not have been president.

Willer (2015), see also Nute (1980) and Fine (2012b)

Similar remarks regarding SDA above also apply to SNCA. We restrict attention to counterfactuals that are interpreted by considering each counterfactual alternative in turn. Thus we do not apply our counterfactual semantics below to the likes of (27).

(27) If you hadn't called both your parents, you would have called your dad.

Unconditionalisation

(28) If Jared had sung it would have been nice, and if Kate had sung it would have been nice.

a. So if Jared or Kate had sung, it would have been nice.

Success

(29) If we had a child then we would have a child.

Weak centering.

(30) K: It's a good thing you're not greedy. If you had eaten all the cookies, there would be none left for Umma.

J: Actually, I ate all the cookies.
K: $\rightsquigarrow$ So there are none left for Umma!

Cumulative transitivity

(31) S: If I were happy, I'd know it.
T: And if you were happy and you knew it, you'd clap your hands!
S: So if I were happy, I would clap my hands.
T: Very good!

We take it that a semantics of counterfactuals ought to validate each of the inference patterns above. But they are not derived rules of the basic imaginative modal logic IK. And so in the following section we introduce a number of axiom schema of imaginative modal logic one can assume in order to derive these desirable inference patterns. The result is a variety of imaginative modal logics that better capture our intuitive patterns of counterfactual inference.

4.3 Six axiom schema for imaginative modal logics

We will not present one imaginative modal logic, but instead present many axiomatisations is a modular fashion, proving soundness and completeness in turn for each axiom scheme with respect to its corresponding semantic property. This is a familiar strategy for modal logicians, and for good reason: on this approach one can pick and mix one's favourite axioms, leaving aside those unsuitable for whatever application is presently at hand.

Table 4.1 below lists each axiom scheme with its corresponding semantic property.[3] Given $\Lambda \subseteq \{$(a), (b), (c), (d), (e), (f)$\}$, the table furnishes 64 imaginative modal logics IK $+ \Lambda$, where IK $+ \Lambda$ is the deductive closure of IK with all instances of the axiom schema in Λ.

Two clarifications about the table are worth mentioning. Firstly, we let $i_w(\varphi)$ be the extension of ι from literals to arbitrary sentences of the base language $\mathcal{L}$ according to the semantics of IE (Definition 3). Secondly, since the semantic properties are given with respect to a single model of IML, for convenience we write $w \models \varphi$ instead of $\mathcal{M}, w \models \varphi$ and leave the subscript $\mathcal{M}$ in $[\![\varphi]\!]_{\mathcal{M}}$ implicit.

Axiom scheme (a) is particularly important because of the following three consequences.

Consequence 1. Antecedent substitution Imaginative equivalences are salva veritate substitutable in counterfactual antecedents. Every model $\mathcal{M}$ of IML validating (a) also validates $(\varphi \equiv \psi) \to [\varphi]\chi \leftrightarrow [\psi]\chi$.

[3] I was first inspired to look for semantic properties such as these while reading Berto (2017).

	Characteristic axiom scheme	Corresponding semantic property
(a)	$(\varphi \hookrightarrow \psi) \to ([\psi]\chi \to [\varphi]\chi)$	*Imaginative inclusion* If $i_w(\varphi) \subseteq i_w(\psi)$ and $wR_\varphi v$ then $wR_\psi v$
(b)	$([\varphi]\chi \wedge [\psi]\chi) \to [\varphi \vee \psi]\chi$	*Unconditionalisation* If $wR_{\varphi\vee\psi}v$ for some $v \in [\![\chi]\!]$, then $wR_\varphi u$ or $wR_\psi u$ for some $u \in [\![\chi]\!]$
(c)	$(\varphi \equiv \psi) \to (\varphi \leftrightarrow \psi)$	*Veracity* If $i_w(\varphi) = i_w(\psi)$ then $w \models \varphi \leftrightarrow \psi$
(d)	$[\varphi]\varphi$	*Success* If $wR_\varphi v$ then $v \models \varphi$
(e)	$[\varphi]\psi \to (\varphi \to \psi)$	*Weak centering* If $w \models \varphi$ then $wR_\varphi w$
(f)	$([\varphi]\psi \wedge [\varphi \wedge \psi]\chi) \to [\varphi]\chi$	*Cumulative transitivity* If $R_\varphi[w] \subseteq [\![\psi]\!]$ and $wR_\varphi v$ then $wR_{\varphi\wedge\psi}v$

Table 4.1: Axiom schema (a)–(f) with their corresponding semantic properties.

However, (a) does not entail the salva veritate substitution of *classical* equivalents in counterfactual antecedents, a rule we found to be undesirable in section 1.2. This has been a longstanding problem for possible worlds analyses of counterfactuals (see Warmbrōd, 1981; Fine, 2012a). Indeed, none of the axiom schema above imply that if $\varphi \leftrightarrow \psi$ is a theorem of classical propositional logic then $[\varphi]\chi \leftrightarrow [\psi]\chi$ is a theorem IK. On the other hand, classical equivalents are salva veritate substitutable in counterfactual *consequents*: any IML proving $\chi \leftrightarrow \delta$ also proves $[\varphi]\chi \leftrightarrow [\varphi]\delta$ by necessitation and K, for any sentences φ, χ and δ of $\mathcal{L}_{IML}$.

Consequence 2. Simplification of disjunctive antecedents (SDA). Since $\varphi \hookrightarrow \varphi \vee \psi$ is a theorem of IE (Fact 19), it is also a theorem of IPL and hence of any normal imaginative modal logic. Thus from (a) we infer $[\varphi \vee \psi]\chi \to [\varphi]\chi$.

Consequence 3. Simplification of negated conjunctive antecedents. This follows from IE's Inclusive De Morgan law (A9). As $\neg\varphi \vee \neg\psi \vee (\neg\varphi \wedge \neg\psi) \equiv \neg(\varphi \vee \psi)$ is a theorem of IE, and hence of any normal imaginative modal logic, by $\equiv$ elimination (Fact 19) $\neg\varphi \vee \neg\psi \vee (\neg\varphi \wedge \neg\psi) \hookrightarrow \neg(\varphi \vee \psi)$ is a theorem of

IK, so $\neg\varphi \hookrightarrow \neg(\varphi \wedge \psi)$ is too by $\hookrightarrow$ simplification (again Fact 19). Then from (a) we infer $[\neg(\varphi \vee \psi)]\chi \rightarrow [\neg\varphi]\chi$.

Now that we have seen some intuitively valid rules of counterfactual inference, one might wonder where to draw the line. In the next section we will highlight one seemingly plausible rule, but which ultimately turns out to be unacceptable. We discuss the rule because its invalidity helpfully illustrates the different semantic natures of counterfactual antecedents and consequents, whereby counterfactual antecedents are interpreted imaginatively but counterfactual consequents truth-conditionally.

4.4 An invalid rule of counterfactual inference

This section is about the following rule of counterfactual inference, which is an axiom of Lewis's basic counterfactual logic V (1981).[4]

$$\frac{[\varphi]\psi \qquad [\psi]\varphi}{[\varphi]\chi \leftrightarrow [\psi]\chi}\ \text{Substitutivity}$$

Substitutivity appears to receive intuitive support.

(32) P1 If Tig came to the party, Hannah would come too.
P2 And if Hannah came to the party, Tig would come too.
P3 And if Tig came to the party, we would have fun.
C So, if Hannah came to the party, we would have fun.

However, a semantics of counterfactuals should not validate substitutivity. We show this by way of counterexample: substitutivity validates some intuitively invalid patterns of inference. To see this, let us first point out that the semantic behaviour of counterfactual consequents appears to be truth-conditional. To illustrate, (33), while hardly assertable, is nonetheless true.

(33) If I had remembered my wallet, I would pay or not pay for dinner.

(33) is true because its consequent, a tautology, is always true. But as we saw in section 1.2, counterfactual antecedents do not behave according to a truth-conditional semantics. We can therefore counterpose the imaginative semantics of counterfactual antecedents with the truth-conditional semantics of counterfactual consequents. In the remainder of this section we will focus on one contrast in particular: the fact that disjunction receives an inclusive semantic interpretation in truth-conditional contexts but a noninclusive semantic interpretation in hypothetical contexts.

[4] Lewis (1971, 80) calls the rule Axiom B. The name *substitutivity* come from Berto (2017, 1291), who discusses it in relation to his notion of imaginative content rather than counterfactuals directly.

Recall the two switches of Ciardelli et al. (2018b), featured in section 1.2.2. In that set-up, the light is on just in case switches A and B are both in the same position (either both up or both down). Suppose both switches are currently up and consider (8b), repeated as (34).

(34) If switch A and switch B were not both up, then switch A or switch B would be down. $[\neg(A \wedge B)](\neg A \vee \neg B)$

Semantically speaking, the *or* of (34)'s consequent receives a truth-conditional, inclusive interpretation, though it may be interpreted exclusively on pragmatic grounds.[5] Nonetheless, granted that truth-conditional disjunction is semantically inclusive, (34) is true.

Now what if substitutivity were a valid rule of counterfactual inference? Then the following argument would be valid, with (34) reappearing as (35b).

(35) a. If switch A or switch B was down, switch A and switch B would not both be up. $[\neg A \vee \neg B]\neg(A \wedge B)$
b. And if switch A and switch B were not both up, switch A or switch B would be down. $[\neg(A \wedge B)](\neg A \vee \neg B)$
c. And if switch A or switch B was down, the light would be off. $[\neg A \vee \neg B]\textsc{Off}$
d. # So, if switch A and switch B were not both up, the light would be off. $[\neg(A \wedge B)]\textsc{Off}$

We encountered (35c) and (35d), respectively as (8a) and (8b) above. Recall that according to the results of Ciardelli et al. (2018b), (35c) is generally judged true but (35d), the argument's conclusion, is generally judged false or indeterminate. Thus substitutivity is not a truth-preserving rule of counterfactual inference.

But if substitutivity is invalid, what are we to make of its intuitive instances? We saw one above in the form of (32). It would seem these are already covered by the rule we called (f) in section 4.3.

$$\frac{[\varphi]\psi \qquad [\varphi \wedge \psi]\chi}{[\varphi]\chi}\ (\mathrm{f})$$

For instance, using (f) we can construct an intuitively valid argument with the same conclusion as (32) above.

[5]The textbook argument establishing that *or* is semantically inclusive goes as follows. If *or* were not semantically inclusive, a speaker of, say, (i) would have contradicted themself.

(i) Julia can sing or dance. In fact, she can do both.

The noninclusive reading of *or* is readily explained by scalar implicature: a more informative alternative utterance to (34) is to assert that if A and B were not both up, both would be down. So a cooperative speaker who opts for the less informative (34) licenses the hearer to reject the stronger assertion, implicating that it A and B were not both up, exactly one would be down.

(36) a. If Hannah came to the party, Tig would come too.
b. And if Hannah and Tig came to the party, it would be fun.
c. So, if Hannah came to the party, it would be fun.

It is likely that many arguments superficially justified by substitutivity can be validly reconstructed using (f), though we will not here check cases in favour of this point. Instead, let us now turn to our main logical result of this chapter: that each axiom scheme is strongly sound and complete with respect to its corresponding class of models.

4.5 Soundness and completeness of each IML

Theorem 39 (Soundness)**.** *For each axiom scheme (a)–(f), every instance of the axiom scheme is true in every model of IML satisfying its corresponding semantic property.*

Proof. A routine checking of cases. Proposition 65 of the appendix. □

Proving completeness of IPL requires a bit more work, but nowhere near the complexity of IPL's completeness proof.

We prove completeness of each imaginative modal logic by suitably adapting the canonical model construction familiar from Lemmon and Scott (1977), Makinson (1966) and Cresswell (1967). First things first, given any normal imaginative modal logic Λ we say a set Γ of sentences of IML is Λ-*consistent* just in case $\perp$ is not derivable from the members of Γ using the rules and axiom schema of Λ. We then define the canonical model for any imaginative normal modal logic like so.

Definition 40 (Canonical model for Λ)**.** *Let Λ be a normal imaginative modal logic. The canonical model for Λ is $\mathcal{M}^\Lambda = (W^\Lambda, \{R^\Lambda_\varphi\}_{\varphi \in \mathcal{L}_{IML}}, V^\Lambda)$, where*

1. *W^Λ is the set of Λ-maximally consistent sets.*
2. *For each R^Λ_φ and $w, v \in W^\Lambda$, we put $wR^\Lambda_\varphi v$ iff $\langle\varphi\rangle\psi \in w$ for every $\psi \in v$.*
3. *V^Λ assigns to each $w \in W^\Lambda$ its canonical IPL model $(\mathcal{I}_w, X_w, \iota_w)$ as in Def. 31.*

For any $V(w) = (\mathcal{I}_w, X_w, \iota_w)$, when $p \in \mathcal{I}_w$ we will often simply write $p \in w$.

The desired completeness result follows from showing that each axiom scheme is canonical for its corresponding semantic property, in the following sense.

Proposition 41. *For any axiom scheme φ, let* $\mathrm{IK} + \varphi$ *be the deductive closure of* IK *with all instances of the axiom scheme φ. Then for $\varphi \in \{$(a), (b), (c), (d), (e), (f)$\}$, the canonical model $\mathcal{M}^\Lambda$ of $\Lambda = \mathrm{IK} + \varphi$ has Λ's corresponding semantic property.*

Proof. We check each imaginative modal logic in turn.

(a) Let $\Lambda = \text{IK} + (\varphi \hookrightarrow \psi) \to ([\psi]\chi \to [\varphi]\chi)$. Pick any $w, v \in W^\Lambda$ such that $\varphi \hookrightarrow \psi \in w$ and $wR^\Lambda_\varphi v$. To show $wR^\Lambda_\psi v$, pick any $[\psi]\chi \in w$. Then as $\varphi \hookrightarrow \psi, [\psi]\chi \in w$, it follows that $[\varphi]\chi \in w$. Then as $wR^\Lambda_\varphi v$, $\chi \in v$.

(b) Let $\Lambda = \text{IK} + ([\varphi]\chi \wedge [\psi]\chi \to [\varphi \vee \psi]\chi)$ and pick any $w \in W^\Lambda$ with $wR^\Lambda_{\varphi\vee\psi} v$ for some $v \in [\![\chi]\!]$. Then $\chi \in v$, so $\langle \varphi \vee \psi \rangle \chi \in w$. And by the above, $(\langle \varphi \vee \psi \rangle \chi \to \langle \varphi \rangle \chi \vee \langle \psi \rangle \chi) \in w$, so by modus ponens $\langle \varphi \rangle \chi \vee \langle \psi \rangle \chi \in w$. As w is a Λ-mcs, $\langle \varphi \rangle \chi \in w$ or $\langle \psi \rangle \chi \in w$. Then by the existence lemma (Lemma 66), $wR^\Lambda_\varphi u$ or $wR^\Lambda_\psi u$ for some $u \in W^\Lambda$ with $\chi \in u$, so $u \in [\![\chi]\!]$.

(c) Immediate.

(d) Let $\Lambda = \text{IK} + [\varphi]\varphi$. Pick any $w, v \in W^\Lambda$ and suppose $wR^\Lambda_\varphi v$, i.e. $\psi \in v$ for every $[\varphi]\psi \in w$. Then as $[\varphi]\varphi \in w$, $\varphi \in v$, as required.

(e) Let $\Lambda = \text{IK} + (\varphi \wedge [\varphi]\psi \to \psi)$. The result is almost immediate. For any $w \in W^\Lambda$ with $\varphi \in w$, and any $[\varphi]\psi \in w$, we have $\psi \in w$, and so $wR^\Lambda_\varphi w$.

(f) Let $\Lambda = \text{IK} + ([\varphi]\psi \wedge [\varphi \wedge \psi]\chi \to [\varphi]\chi)$. Pick any $\varphi \in \mathcal{L}$, $\psi \in \mathcal{L}_{CFL}$ and $w, v \in W^\Lambda$. Suppose $R^\Lambda_\varphi \subseteq [\![\psi]\!]$ and $wR^\Lambda_\varphi v$.

We first show that $[\varphi]\psi \in w$. For suppose not. Then as w is a Λ-mcs, $\neg[\varphi]\psi \in w$, so by dual, $\langle \varphi \rangle \neg\psi \in w$. Then by the existence lemma (Lemma 66 of the appendix), for some $u \in W^\Lambda$ we have $wR^\Lambda_\varphi u$ and $\psi \notin u$, contradicting the fact that $R_\varphi[w] \subseteq [\![\psi]\!]$. Thus $[\varphi]\psi \in w$. We have to show that $wR^\Lambda_{\varphi\wedge\psi} v$. To that end, pick any $[\varphi \wedge \psi]\chi \in w$. Then as $[\varphi]\psi, [\varphi \wedge \psi]\chi \in w$, by modus ponens, $[\varphi]\chi \in w$. And since $wR^\Lambda_\varphi v$, $\chi \in v$, as desired.

□

We now have everything in place to prove completeness. First, some usual notation: for any normal imaginative modal logic Λ, sentence φ and set of sentences Γ of IML, write $\Gamma \vdash_\Lambda \varphi$ just in case φ is deducible from Γ using the rules and instances of the axiom schema of Λ, and write $\Gamma \models_\Lambda \varphi$ just in case every model of Λ making every member of Γ true also makes φ true.

Theorem 42 (Strong completeness of each IML)**.** *Let φ be a sentence and Γ a set of sentences of IML, and $\Lambda = \text{IK} + \varphi$ where φ is one of (a–f). If $\Gamma \models_\Lambda \varphi$ then $\Gamma \vdash_\Lambda \varphi$.*

Proof. We use the results of Blackburn, de Rijke, and Venema (2002, chapter 4).

Contrapositively, suppose $\Gamma \nvdash_\Lambda \varphi$. Then $\Gamma \cup \{\neg\varphi\}$ is Λ-consistent, and so by Lindenbaum's lemma extendable to a Λ-mcs $w \in W^\Lambda$. By Proposition 41, the canonical model for Λ (Def. 40) is indeed a model of Λ. And as $\Gamma \cup \{\neg\varphi\} \subseteq w$, $\mathcal{M}^\Lambda, w \models \Gamma$ but $\mathcal{M}^\Lambda, w \models \neg\varphi$, and so $\mathcal{M}^\Lambda, w \not\models \varphi$. Therefore, $\Gamma \not\models \varphi$. □

4.6 Adding time

There is one sense in which the above models of imaginative modal logic are not yet rich enough to serve as models of counterfactual reasoning. The missing ingredient is time. So often when we evaluate a counterfactual we allow some time to pass between the truth of the antecedent and the consequent, as in (37).

(37) If you hadn't sat beside me in class, we wouldn't be married today.

We would like our models to be able to interpret sentences such as (37), but the worlds of IML models defined above are static, in the sense that they assign to each atomic sentence a fixed truth value. Our worlds might therefore be better described as *moments*. And so our semantics above seems to assume that when we imagine a counterfactual antecedent true, its consequent must hold immediately.

However, the observation above—that in evaluating a counterfactual we allow some time to pass between the truth of the antecedent—does not apply universally, as (38) shows.

(38) ?? If it were the weekend, then it would be a weekday.

A weekday always succeeds the weekend in time, but that might not stop one objecting to (38) on the grounds that, if it were the weekend, it would very well be the weekend—and not a weekday. Still, one could retort that, if it were the weekend, it would *eventually* be a weekday (see 6.2 for further discussion).

The status of tense in counterfactuals raises a unique set of complex issues, ones we will not consider here (though see e.g. Iatridou, 2000; Schulz, 2008, 2017; Ippolito, 2013). But to have a greater semblance of intuitive plausibility, we would like our models to feature a temporal dimension. And since a temporal dimension is easy to add, we will add it here. To that end, given a set W of worlds, let us consider an order $\to \subseteq W \times W$, representing *lawful temporal succession*. The intended interpretation of the relation is that $w \to v$ just in case it is possible for moment v to succeed moment w after one step in time. Thus $\to$ encodes a notion of physical modality via temporal modality, describing which moments may succeed which in time.

It is straightforward to add a dynamic twist to IML's model construction.

Definition 43 (Dynamic IML model)**.** *A dynamic IML model is pair $(\mathcal{M}, \to)$ where,*

1. *$\mathcal{M}$ is an IML model with set of worlds W.*

2. *$\to \subseteq W \times W$ is a binary relation over W.*

We say a path is a sequence of worlds $(w_1, w_2, \dots)$, where $w_i \to w_{i+1}$ for each $i \geq 1$. And we call a path terminal just in case it is infinite or its final state has no successors.

Using the temporal succession relation we may introduce the following two temporal operators.

Definition 44 (Temporal modals)**.** *We add to the language of IML the operators $\Box$ and $\Diamond$, with the semantic clauses below, for any dynamic IML model $\mathcal{M}$ and world w.*

$$\begin{array}{lll} \mathcal{M}, w \models \Box\varphi & \textit{iff} & \textit{every terminal path beginning with } w \\ & & \quad \textit{contains a world } v \textit{ with } \mathcal{M}, v \models \varphi. \\ \mathcal{M}, w \models \Diamond\varphi & \textit{iff} & \textit{some path beginning with } w \\ & & \quad \textit{contains a world } v \textit{ with } \mathcal{M}, v \models \varphi. \end{array}$$

The sentence $\Box\varphi$ intuitively states that no matter what sequence of moments follow from the present moment, φ will be true at some point, that is, φ *must* occur in future. Dually, the sentence $\Diamond\varphi$ states intuitively that from the present moment, φ *might* be true in future.[6]

4.6.1 Blending imaginative and temporal modalities

With time in the picture, we can suggest the following semantic clause for counterfactuals, bearing in mind that according to it, for example, (38) comes out true. And so we call the clause below a semantics of an 'open-ended' interpretation of counterfactuals.

Definition 45 (Open-ended counterfactuals)**.** *Let $(\mathcal{M}, \rightarrow)$ be a dynamic IML model with world w. We give the following semantic clause for the binary connective $>$.*

$$\mathcal{M}, w \models \varphi > \psi \quad \textit{iff} \quad \mathcal{M}, w \models [\varphi]\Box\psi$$

An open-ended interpretation of counterfactuals is characterised by having no temporal bound on when we check for the truth of the consequent after imagining the antecedent true. In other words, an open-ended counterfactual is true at a world just in case according to every way of imagining the antecedent true, the consequent will eventually hold.

4.6.2 Adding time rather than causality

Let us briefly pause to address one concern about adding time to our models of counterfactual reasoning. Given the tight psychological connection between counterfactual and causal reasoning (see Byrne, 2007, chapter 5), a promising alternative proposal would be to add a causal rather than temporal component to our models (for example Galles and Pearl, 1998; Halpern, 2000; Hiddleston,

[6]These temporal modalities are taken from the much more expressive computation tree logic (CTL) of Clarke and Emerson (1982). Note that in CTL, $\Box$ is referred to as AF and $\Diamond$ as EF.

2005; Schulz, 2011; Briggs, 2012, develop a causal modelling semantics for counterfactuals). However, the suggestion that the semantics of counterfactuals requires a causal component is ruled out by counterfactuals with trivially true consequents, as in (39).

(39) If it were Tuesday, two plus two would be four.

Although (39) is completely uninformative, it is nonetheless true. But where is causality? The days of the week do not cause the truths of arithmetic. There is no causal relationship between the antecedent and consequent of a counterfactual whose consequent is trivially true, as in (39).[7]

In contrast, the purely temporal modalities of Definition 44 can account for the truth of (39). To see this, note that $\Box\varphi$ is a logical consequence of φ; for, every terminal path beginning with a world w making φ true contains a world (namely, w itself) making φ true. Thus, whatever way the world w would be if it were Tuesday, since every world makes true $2 + 2 = 4$, by the above every world also makes true $\Box(2 + 2 = 4)$. Therefore, whatever way the actual world is, it makes true $[\text{Tuesday}]\Box(2 + 2 = 4)$, that is, Tuesday $> 2 + 2 = 4$. And so we derive the truth of (39) using our purely temporal modalities, as desired.

This, then, is the general framework for a counterfactual logic understanding counterfactual antecedents in terms of their imaginative content. But from this bird's eye view, one might fear we have traded depth for scope. After all, there is a restless quality to the approach of this chapter: adding axiom schema and semantic conditions piecemeal. For how many individual inference patterns must we validate before we can say we are done? I suggest we stop sketching and instead, by analysing the concepts within, aim to reproduce the validities of counterfactual inference from a unified standpoint.

[7] Nevertheless, if one were to assert (39) the listener would be licensed to conclude the speaker is attempting to establish some causal connection between the calendar and arithmetic. This inference is naturally explained on pragmatic grounds, namely, by scalar implicature. Consider (i), a more informative alternative utterance to (39).

(i) If it were Tuesday, two plus two would be four, and also, if it were not Tuesday, two plus two would be four.

A listener who only hears (39) would be licensed to conclude that the speaker is not in a position to assert the more informative (i), implicating that according to the speaker, if it were not Tuesday, two plus two might not be four. In other words, the speaker implicates some difference-making relationship between the day being Tuesday and two plus two being four. Just such a difference-making condition is a key clause in a number of analyses of causation (Lewis, 1973a; Braham and Van Hees, 2012, Definition 3). We may thus explain the causal flavour of (39) as the implicature of a difference-making condition, which is plausibly itself a causal condition.

The central primitives of the model theory of normal imaginative modal logic are the accessibility relations. We have talked at length about the ways of imagining a sentence true, with each accessibility relation R_φ representing the 'ways of imagining' φ true at a given world. But what are these 'ways of imagining', really? Are they irreducibly primitive, or can we construct them from simpler, more concrete parts? Our talk of ways of imagining is at the heart of the present account. It is reasonable, then, to assert that the capacity of normal imaginative modal logics to represent our intuitive counterfactual reasoning rests on our ability to explain what these 'ways of imagining' intuitively are.

Thus our final substantive chapter is devoted to analysing what ways of imagining are, without recourse to opaque primitives. We will take a way of imagining to be *a list of atomic changes to the actual world*. This interpretation will result in a class of models of normal imaginative modal logic and pave the way for our last main result, that these models—the 'standard models', it is tempting to say—are each models of the desirable axiom schema (a)–(f) we encountered in this chapter.

Chapter 5

A change of world

Our imagination often exhibits a surgical ability to create fictitious scenarios from the actual one. Just imagine the chair beneath you disappeared. Or these words were printed blue. Each scenario comes easily to mind. As long as we know the meaning of a sentence, it is not so hard to consider actual world changed to imagine the sentence true.

This imaginative act is a basic feature of how we interpret counterfactuals. Counterfactuals and imaginative content are united by the simple thought that in evaluating a counterfactual, we imagine its antecedent true. This shift to a hypothetical context, this change of world, is an act of imagination. On the present line of thought, one very direct way to evaluate, say, the counterfactual, "If the chair beneath you disappeared, you would fall down" is to imagine the actual world changed with the chair gone and simulate what happens. In a slogan, we imagine a change, press play, and check for the consequent.

Now, not all changes are so obvious to imagine. Things are more interesting when we investigate counterfactuals with logically complex antecedents. Consider, for instance, counterfactuals with disjunctive antecedents like (40), which as Alonso-Ovalle (2006, 2009) discusses, have long been a thorn in the side of the minimal change semantics of Lewis (1973b) and Stalnaker (1968).

(40) If we had had good weather this summer or the sun had grown cold, we would have had a bumper crop.

(Adapted from Nute, 1975)

What does it mean to imagine the actual world changed to make us "have good weather this summer or the sun grow cold"? This complex antecedent seems not to invite a single, 'disjunctive' change but two simpler changes, one for each disjunct. Our semantics of imaginative content captures this fact by assigning two imaginative contents to the disjunction; namely, the imaginative content of having good weather this summer and of the sun growing cold. And on the approach of imaginative content pursued in this essay, each 'simple change' is in turn described by the imaginative content of atomic sentences and negations of atomic sentences.

This suggests that counterfactual interpretation happens in two stages.

1. **Atomise the antecedent.** Decompose the antecedent until it is described by sets of imaginative contents of literals.

2. **Check for the consequent.** Check whether the consequent holds under each way of imagining the actual world changed by these imaginative contents.

Counterfactual semantics is thus a blend of imaginative and informative content, as we saw in the previous chapter. We *imagine* the antecedent and check for the *truth* of the consequent. And so a semantics of counterfactuals needs a model uniting both kinds of content. But there are some preliminary concerns to address about how these two kinds of content are to interact, which we address now.

5.1 Subjective and objective ways of imagining

A person's imagination can say a lot about their personality. A child's play with dolls, say, can often speak to their home life. In general two people with the same prompt will often bring very different scenarios to mind. On the other hand, counterfactuals often have a determinate connection to reality: we take the more sober counterfactuals to have determinate truth values. For instance, we use counterfactuals to justify praise and blame, arguing in court that a harm would been avoided but for a defendant's actions.

The subjectivity of imagination would seem to pose a problem for a semantics of counterfactuals in terms of imaginative content. It is a legitimate question how counterfactuals, which are often so worldly, can be understood in terms of something as subjective as imaginative content.

To illustrate, a colour blind person might imagine red and green as the same colour in the mind's eye. What we have said so far would seem to suggest that according to such a colour blind person (41a) and (41b) should have the same truth value.

(41) a. If the traffic light were red, it would be red.
     b. If the traffic light were green, it would be red.

But a colour blind person would likely recognise that their imagination is not veridical when it comes to distinguishing red and green, and so would take (41a) and (41b) to have different truth values.

The subjective character of the imagination appears to conflict with the determinate truth of many counterfactuals. We want the imagination of counterfactual antecedents to still speak about imagination, while also being sufficiently objective, conforming to actuality. But what exactly does it mean to demand that one's imagination conform to actuality?

We bridge the subjective–objective divide by having imaginative equivalence entail material equivalence. This is the principle we called *veracity* above. Veracity guarantees, for instance, that we cannot take 'the light is red' and 'the light is green' to be imaginatively equivalent in any world where the light is red but not green, or green but not red. The principle of veracity also ensures, for example, that imagining John a bachelor means imagining him an unmarried man, imagining someone innocent cannot mean imagining them guilty, and so on.

We would nonetheless like to retain some looseness in the counterfactual imagination. After all, not all counterfactuals have a determinate truth value. Goodman's Caesar pair is the textbook example.

(42) a. If Caesar were in command, he would use the atom bomb.
b. If Caesar were in command, he would use catapults.
(*op. cit.* Quine, 2013, 203)

Our model will decide the imaginative content of the atomic sentence "Caesar is in command". In so doing, the model will also decide whether it is the same as the imaginative content of "Caesar is in command using the atom bomb" and of "Caesar is in command using catapults"; that is, whether ι(Caesar is in command) is identical to ι(Caesar is in command using the atom bomb), to ι(Caesar is in command using catapults), or to neither. These are not questions for a logician to decide, so we incorporate the looseness into the model construction itself.

5.2 Model construction

We are now in a position to define a model combining (i) IE's treatment of imaginative content with (ii) the treatment of informative content provided by possible worlds and (iii) a temporal dimension discussed in section 4.6 above. Our model of counterfactuals will not feature any imaginative accessibility relations: we will not introduce them as black-box parameters of the model itself but instead generate them from the following model construction (Definition 46).

Definition 46. *A counterfactual model is a tuple* $(W, \to, X, \iota)$*, where*

1. (X, ι) *is a model of IE.*
2. *Each* $w \in W$ *is a model of propositional logic (set of atomic sentences) satisfying veracity: for any* $\varphi, \psi \in \mathcal{L}$*, if* $(X, \iota) \models \varphi \equiv \psi$ *then* $w \models \varphi \leftrightarrow \psi$*.*
3. $\to \subseteq W \times W$ *is a binary relation (representing lawful temporal succession).*

Some notes on the definition of a counterfactual model are in order. Firstly, instead of taking each model w of propositional logic to be an interpretation function from the set of atomic sentences to the truth values $\{0, 1\}$, we take it to simply be the set of atomic sentences true at w. We adopt this approach because it allows the familiar set theoretic operations such as union and inclusion to seamlessly apply when comparing models of propositional logic. So we put $w \models p$ just in case $p \in w$ (and hence $w \models \neg p$ just in case $p \notin w$).

Secondly, for simplicity we will take each world to share the same model of IE; that is, the same assignment of imaginative content to literal sentences. This will make proofs of facts involving counterfactual models considerably easier. It does, however, hold us back from considering changes to the underlying model of IE itself, as we are asked to do when, say, we hear the antecedent, "If to imagine John a bachelor were to imagine him a married man, ...". The reader is very welcome to consider extending the present approach to such antecedents, though we will here only consider counterfactual antecedents expressible in the base language $\mathcal{L}$.

With our definition of model in place, we can now turn to analysing our accessibility relations in terms of lists of atomic changes to the actual world.

5.3 A counterfactual semantics of atomic change

Consider the moment in which I am sitting on my chair. We can represent the fact of the chair being beneath me using the atomic sentence b and by letting w denote the actual world. Then the actual world makes b true; in our formalism, $b \in w$. Now what if my chair suddenly disappeared?

The chair's presence beneath me appears to be a simple fact, one I can switch on and off in my head at will. So I suggest that to model such a simple act of the imagination, we need but simple tools. Formally, we remove the fact of the chair being beneath me from the actual moment. Desiring to have w make $\neg b$ true, we could simply take $w - \{b\}$, for as $b \notin w - \{b\}$, we have $w - \{b\} \models \neg b$. But this is a bit too simple; or better, it is a bit too syntactic to fit a semantic treatment. For why should the use of that particular atomic b matter? Suppose we had a distinct atomic sentence s representing the sentence, 'The seat is beneath me', rather than 'The chair is beneath me' (b). The atomics b and

s may represent distinct sentences, but clearly they share the same imaginative content: $\iota(b) = \iota(s)$. We might say that b and s are just symbols to get at their imaginative content, which is the real medium by which we imagine changes to actuality. So to a more systematic—*viz.* more semantic—end, in imagining the chair beneath me gone, $\neg b$, we will take the set $\{p : \iota(\neg p) = \iota(\neg b)\}$ of atomic sentences whose negation is identical to the imaginative content of $\neg b$, and remove each of those from the actual world, resulting in

$$w - \{p : \iota(\neg p) = \iota(\neg b)\}.$$

As $s \notin w - \{p : \iota(\neg p) = \iota(\neg b)\}$, the resulting world makes s false, as desired.

Of course in imagining the actual world changed we often make atomics true. For example, where f is an atomic sentence representing the floor being soft, we may consider imagining the floor soft. To achieve this in a semantic way, we consider the set $\{p : \iota(p) = \iota(f)\}$ of atomics with the same imaginative content as f and add this to the actual world with the chair removed:

$$\big(w - \{p : \iota(\neg p) = \iota(\neg b)\}\big) \cup \{p : \iota(p) = \iota(f)\}.$$

In general, let us define the positive and negative changes given by a set of imaginative contents as follows.[1]

Definition 47 (Atomic changes)**.** *Let $(W, \rightarrow, X, \iota)$ be a counterfactual model and $c \subseteq X$ a nonempty set of contents. Where At is the set of atomic sentences, we define the positive and negative changes of c, respectively, like so.*

$$c^+ = \{p \in At : \iota(p) \in c\}$$
$$c^- = \{p \in At : \iota(\neg p) \in c\}$$

And for any $w \in W$, we define w^c, called w under the change of c, as follows.

$$w^c = (w - c^-) \cup c^+$$

Thus the positive change of c is the set of atomic sentences whose imaginative content is in c, and the negative change is the set of atomic sentences whose negation's imaginative content is in c.

Now that we have the changes, we can apply them to our possible worlds. But before we do so, one point about our use of possible worlds is in order.

[1] This discussion of changes is inspired by recent interventionist approaches to counterfactuals (e.g. Schulz, 2011; Briggs, 2012; Santorio, 2016), though these approaches do not make use of the present notion of imaginative content nor the modal framework.

5.3.1 Counterpossibles and consistency

Since we are working with possible worlds, we will only consider acts of imagination that result in imagining a possibility. That is to say, the present framework will adopt a vacuist approach to counterpossibles (counterfactuals with impossible antecedents), according to which all counterfactuals with an impossible antecedent are vacuously true. There is an ongoing debate whether some counterpossibles are false (see Brogaard and Salerno, forthcoming; Williamson, 2017; Berto et al., 2018). We adopt a vacuist position for two reasons, one technical and one philosophical. The technical reason is that a restriction to possible worlds, leaving out any impossible ones, allows our approach to fit within the model theory of modal logic, which is built on classical foundations. Though the atomic change approach can likely be enriched straightforwardly with impossible worlds—a task we leave to the interested reader.

The philosophical reason is that counterfactuals appear to be ordinarily evaluated according to a presupposition that in imagining the antecedent true one imagines a possibility. To take a well-worn example, despite the fact that squaring the circle is an impossible geometric construction, many judge (43) false.

(43) If Hobbes had squared the circle, sick children in the mountains of South America at the time would have cared.

(Adapted from Nolan, 1997, 544)

On the present imaginative framework, one can explain why (43) sounds false in terms of imaginative contents. For someone who feels (43) is false might have unwittingly assigned the imaginative content of, say, "Hobbes proved a geometric result" to "Hobbes squared the circle". In this vein, Williamson suggests that in imagining Hobbes squaring the circle one might actually imagine Hobbes "doing geometry in the secrecy of his room" (2018, 364). Philosophers arguing that some counterpossibles, like (43), are false have yet to address the proposal that all counterpossibles are in fact true but can be *believed* false due to an erroneous subjective assignment of imaginative contents.

As a consequence of the fact that we only consider possible worlds, we will restrict attention to consistent imaginative contents; that is, imaginative contents c such that for no atomic sentence p does c contain both $\iota(p)$ and $\iota(\neg p)$. Recall from the semantics of CIE (Definition 22) that for any model (X, ι) of IE and $\varphi \in \mathcal{L}$,

$$ci(\varphi) = \{c \in i(\varphi) : \text{for no atomic sentence } p \text{ are both } \iota(p), \iota(\neg p) \in c\}.$$

The following fact is then an immediate consequence of the definition of $ci(\varphi)$.

Fact 48. *For any $\varphi \in \mathcal{L}$ and $c \in ci(\varphi)$, c^+ and c^- are disjoint. Hence the order of constructing w^c does not matter: $(w - c^-) \cup c^+ = (w \cup c^+) - c^-$ for any world w.*

5.3.2 Constructing the imaginative accessibility relations

Now back to applying our atomic changes to possible worlds. The object w^c above is intended to represent the world we imagine when we change w to feature the imaginative content in c. But what does this have to do with counterfactual antecedents? Recall that according to the semantics of IE, the imaginative content of each sentence $\varphi \in \mathcal{L}$ is a set $i(\varphi)$ of such changes. Our definition of atomic changes above, then, gives an exact meaning to the idea that $i(\varphi)$ contains the 'ways of imagining φ true', and $ci(\varphi)$ the 'consistent ways of imagining φ true.' Given any world w, $ci(\varphi)$ contains a list of ways to consistently change w to imagine φ true.

This line of thought can be naturally phrased in terms of modal logic. For any counterfactual antecedent $\varphi \in \mathcal{L}$ we construct an imaginative accessibility relation R_φ whereby v is R_φ-accessible from w just in case v is the result of changing w according to a way of imagining φ true, that is, v is one way to imagine φ holding at w.

Definition 49 (Atomic change model)**.** *An atomic change model is a quadruple* $(W, \to, \{R_\varphi\}_{\varphi\in\mathcal{L}}, X, \iota)$ *where*

1. $(W, \to, X, \iota)$ *is a counterfactual model (Definition 46).*
2. $\{R_\varphi\}_{\varphi\in\mathcal{L}}$ *is a set of relations, one for each sentence in the base language* $\mathcal{L}$*, where for any* $w, v \in W$*,*

$$wR_\varphi v \quad \textit{iff} \quad v = w^c \textit{ for some } c \in ci(\varphi).$$

Note that each counterfactual model uniquely determines its atomic change model. In this sense we may claim to have analysed the primitives of the model theory of IML purely in terms of possible worlds and imaginative content.

Let's see the atomic change approach to counterfactuals in action by looking at a simple example.

5.4 Worked example: antecedent strengthening

Figure 5.1 depicts a scenario with an evil barista. Say w is the actual world, where, for simplicity, we have every atomic sentence false at w, and so $w = \emptyset$. In particular, at w I am not drinking coffee. Now consider the world v in which every atomic is false, expect for 'coffee', representing the fact that I have coffee: $v = \{\text{coffee}\}$. Let ☕ be the imaginative content of me drinking coffee: $\iota(\text{coffee}) = \text{☕}$, and let $\iota(l) \neq \iota(l')$ for all distinct literals l, l'. Then $\{\text{☕}\}^+ = \{\text{coffee}\}$, $\{\text{☕}\}^- = \emptyset$, and so $w^{\{\text{☕}\}} = (w - \emptyset) \cup \{\text{coffee}\} = v$. Hence $wR_{\text{coffee}}v$.

Similarly, let $u = \{\text{coffee, poison}\}$ be a world in which I am drinking coffee and poison, with $\iota(\text{poison}) =$ ☠. Then $i(\text{coffee} \wedge \text{poison}) = \{\{$☕, ☠$\}\}$, and $\{$☕, ☠$\}^+ = \{\text{coffee, poison}\}$. So $u = w^{\{☕, ☠\}}$, that is, $wR_{\text{coffee}\wedge\text{ poison}}u$.

Regarding dynamics, suppose if I am only drinking coffee, I become happy; if I am drinking coffee and poison, I become sick.

$$\{\text{coffee}\} \to \{\text{happy}\}$$
$$\{\text{coffee, poison}\} \to \{\text{sick}\}$$

Then we have $\mathcal{M}, w \models \text{coffee} > \text{happy}$ but $\mathcal{M}, w \not\models \text{coffee} \wedge \text{poison} > \text{happy}$, since w is related by $R_{\text{coffee}\wedge\text{poison}}$ to a world where I am not happy. As desired, then, at w it is true that if I were drinking coffee, I would be happy, and it is false that I were drinking coffee and drinking poison, I would be happy.

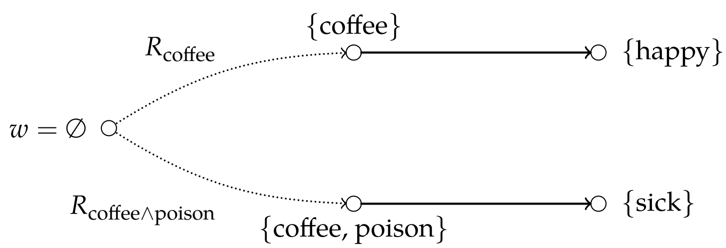

Figure 5.1: Countermodel to antecedent strengthening

The toy model above serves as a countermodel to antecedent strengthening, the undesirable inference from $\varphi > \chi$ to $\varphi \wedge \psi > \chi$. The fact that antecedent strengthening is invalid on the similarity approaches of Lewis (1973b) and Stalnaker (1968) was viewed as a major advantage of the similarity approach over rival counterfactual analyses in terms of material and strict implication, which do validate antecedent strengthening. Happily, our example of Figure 5.1 shows how the atomic change semantics of counterfactuals also swiftly avoids the problem.

We turn now to some inferences that the atomic change semantics validates.

5.5 Validities of the atomic change semantics

This section is devoted to proving the following result.

Theorem 50. *The axiom schema (a)–(f) are true in every atomic change model.*

Proof. We show that every atomic change model satisfies the semantic property corresponding to each axiom scheme. The result then follows from Theorem 65.

Pick any atomic change model $\mathcal{M} = (W, \{R_\varphi\}_{\varphi \in \mathcal{L}}, X, \iota)$. Since $\mathcal{M}$ is clear, to save ink we write $w \models \varphi$ instead of $\mathcal{M}, w \models \varphi$. Imaginative inclusion is immediate since $i(\varphi \vee \psi) = i(\varphi) \cup i(\psi)$ by the semantics of IE. Unconditionalisation is also immediate since $w \models \varphi \hookrightarrow \psi$ iff $i(\varphi) \subseteq i(\psi)$ by definition. Veracity holds by assumption on counterfactual models (Definition 46).

Success. Firstly let $\mathbb{L}$ be the set of literals and $\leq$ the fixed linear order over the literals used in the construction of φ's standard disjunctive normal form φ^*. Given any atomic change model $\mathcal{M}$ and finite set of contents $c \subseteq X$, we define the *standard literal form* of c (with respect to $\mathcal{M}$), denoted c^ℓ, as follows

$$c^\ell = \bigwedge \{l \in \mathbb{L} : \iota(l) \in c \text{ and } l \leq l' \text{ for every literal } l' \text{ with } \iota(l) = \iota(l')\}.$$

The idea is that c^ℓ is a sentence conjoining a representative literal for each element of c. Note that c^ℓ is well-formed as c, by semantic definition (Definition 3), is finite. Now pick any $\varphi \in \mathcal{L}$ and suppose $wR_\varphi v$, i.e. $v = (w - c^-) \cup c^+$ for some $c \in ci(\varphi)$.

We show that $v \models c^\ell$ by showing that $v \models l$ for every literal l of c^ℓ. Pick any conjunct l of c^ℓ. If l is an atomic sentence p then $\iota(p) \in c$ by definition of c^ℓ, and so $p \in c^+$. Then as $v = (w - c^-) \cup c^+$, we have $c^+ \subseteq v$, so $p \in v$, i.e. $v \models p$. And if l is $\neg p$ for some atomic sentence p then as $\neg p$ is a literal of l, $\iota(\neg p) \in c$, and so $p \in c^-$. Then as $c \in ci(\varphi)$, by Fact 48 $\iota(p) \notin c$ so $p \notin c^+$. Thus $p \notin v$, i.e. $v \not\models p$, and so by classical logic $v \models \neg p$. Either way, then, $v \models c^\ell$.

Now by construction of φ's standard disjunctive normal form φ^* and Lemma 12, we may write φ^* as $\bigvee_{c \in i(\varphi)} c^\ell$. So as $v \models c^\ell$, and $ci(\varphi) \subseteq i(\varphi)$, $v \models c^\ell$ for some $c \in i(\varphi)$, so $v \models \varphi^*$. And by the standard normal form theorem, $v \models \varphi^* \equiv \varphi$, so by veracity, $v \models \varphi^* \leftrightarrow \varphi$, and so $v \models \varphi$, as required.

Weak centering. Suppose $w \models \varphi$. Then by the standard disjunct normal form theorem (Lemma 9) and veracity, $w \models c^\ell$ for some $c \in i(\varphi)$. And as w is a model of propositional logic, c must be consistent, i.e. $c \in ci(\varphi)$. We now show that $wR_\varphi w$, that is, $w = (w - c^-) \cup c^+$.

For the left-to-right inclusion, pick any $p \in w$. First suppose toward a contradiction that $p \in c^-$. Then $\iota(\neg p) \in c$, so as $w \models c^\ell$, $w \models l$ for some literal l with $\iota(l) = \iota(\neg p)$. Then by veracity $w \models l \leftrightarrow \neg p$, so $w \models \neg p$, i.e. $p \notin w$, contradicting our original assumption. Then as $p \in (w - c^-)$, a fortiori $p \in (w - c^-) \cup c^+$, as required.

Conversely, pick any $p \in (w - c^-) \cup c^+$. We show $p \in w$. If $p \in w - c^-$ then $p \in w$ and we are done. So suppose $p \in c^+$. Then $\iota(p) \in c$, so as $w \models c^\ell$, $w \models l$ for some literal l with $\iota(l) = \iota(p)$. By veracity, $w \models l \leftrightarrow p$, so $w \models p$, i.e. $p \in w$.

Cumulative transitivity. We prove the stronger result that for any $w, v \in W$, if $wR_\varphi v$ and $v \models \psi$ then $wR_{\varphi \wedge \psi} v$.

By the semantics of IE, $i(\varphi \wedge \psi) = i(\varphi) \Cup i(\psi)$, so we have to show that $v = \big(w - (x \cup y)^-\big) \cup (x \cup y)^+$ for some $x \in i(\varphi)$ and $y \in i(\psi)$. Now, as $wR_\varphi v$,

$v = (w - c^-) \cup c^+$ for some $c \in i(\varphi)$. We will take $x = c$. Now, by the standard disjunctive normal form theorem, $v \models \psi \equiv \psi^*$, and so $v \models \psi^* \leftrightarrow \psi$ by veracity. Then as $v \models \psi$, we have $v \models \psi^*$. Recall from the proof of *success* above that ψ can be written as $\bigvee_{d \in i(\psi)} d^\ell$. Then by classical disjunction, $v \models d^\ell$ for some $d \in i(\psi)$. And as v is a model of propositional logic, d must be consistent, so $d \in ci(\psi)$. We will take $y = d$. That is, we show that for every atomic p,

$$p \in v \text{ iff } \big(p \in w \text{ and } \iota(\neg p) \notin c \cup d\big) \text{ or } \iota(p) \in c \cup d.$$

For the left-to-right direction, pick any atomic sentence $p \in v$. Then as $v = (w - c^-) \cup c^+$, either $p \in w$ and $\iota(\neg p) \notin c$, or $\iota(p) \in c$. If the latter, $\iota(p) \in c \cup d$ and we are done. So suppose $p \in w$ and $\iota(\neg p) \notin c$. It remains to show that $\iota(\neg p) \notin d$. Suppose for reductio that $\iota(\neg p) \in d$. Then as $v \models d^\ell$, by construction of d^ℓ we have $v \models l$ for some literal l with $\iota(l) \in c$ and $\iota(l) = \iota(\neg p)$. Then by veracity, $v \models l \equiv \neg p$, and so $v \models \neg p$. By classical logic, $v \not\models p$, i.e. $p \notin v$, contradicting the fact that $p \in v$.

For the right-to-left direction, first suppose $\iota(p) \in c \cup d$. If $\iota(p) \in c$ then as $c^+ \subseteq v$, we have $p \in v$. So suppose $\iota(p) \in d$. Then as $v \models d^\ell$, by a similar argument as above, $v \models p$, i.e. $p \in v$, as required. Now suppose $p \in w$ and $\iota(\neg p) \notin c \cup d$. A fortiori, $\iota(\neg p) \notin c$, so $p \in w - c^-$. Then as $wR_\varphi v$, we have $w - c^- \subseteq v$, so $p \in v$ and we are done. □

It is remarkable that the atomic change model—a model constructed according to a particular intuition of atomic change, with no regard for general validities—would turn out to be a model of the desirable axiom schema (a)–(f). When constructing the atomic change model, I did not set out to create a model of those schema. But surprisingly, that is what we have ended up with.

Even though the atomic change model validates many desirable inference patters of counterfactuals, it is far too soon to move from this fact to the idea that the atomic change model is the one correct model of our counterfactual reasoning. For one thing, our result does not make any claims to uniqueness: many other intuitive approaches to counterfactuals may well satisfy the intuitive inference patterns above.

And another thing, our intuitive interpretation of counterfactuals seems to involve more parameters than the atomic change semantics provides. The following section considers one such parameter: our use of *implicit* hypothetical contexts when we interpret counterfactuals.

5.6 Implicit hypotheticals

I would like to avoid the impression that the atomic change semantics by itself offers a comprehensive semantics of counterfactuals. It does not. This section is devoted to one feature of our counterfactual imagination: our openness to considering hypothetical contexts beyond those explicitly raised by a counterfactual antecedent. For it seems the imagination is highly suggestible, open to bringing extra hypothetical contexts to bear on counterfactuals that do not explicitly mention them.[2]

As an illustration, consider the following dialogue between Jenny and Bob.

(44) [JENNY *flips a coin and hides it in her hands.*]

J: Heads or tails?

B: Heads.

J: [*Reveals the coin.*] It's heads: you win! Well done!

(45) **Ending 1.**

B: Yay! If I had bet tails, I would have lost.

We have a strong feeling that what Bob said is true: if Bob had bet tails he would have lost. But consider the following alternate ending to their dialogue.

(46) **Ending 2.**

B: Well, it was all down to luck. The coin could have just as easily landed tails.

J: Yes, but you could have just as easily bet tails.

B: True. If I had bet tails, well, the coin might have landed tails and I would have won anyway.

J: Right! So it's false to say: if you had bet tails, you would have lost. For if you had bet tails, you might have still won!

In the dialogue with the first ending, it seems we fix the fact that the coin landed heads. But in the dialogue with the second ending, Bob invites us to stretch our imagination by unfixing the fact that the coin landed heads, a suggestion Jenny goes along with. Now, the truth of (47) all depends on whether we fix the result of the coin flip.

(47) If Bob had bet tails, he would have lost.

Fixing that the coin landed heads, (47) is true. But allowing that the coin might have landed tails, if Bob had bet tails then still the coin might have landed tails, and Bob would have won, making (47) false. So what is the truth of the matter?

I do not think there is one. Each truth value of (47) seems appropriate to the ending in which it is uttered. Nonetheless, default interpretations can act in

[2] This section was inspired in particular by the discussion of a background parameter by Ciardelli et al. (2018b, §4), though as their discussion makes use of inquisitive propositions, their account undesirably makes semantically equivalent some counterfactuals that, as we saw in section 1.2, ought to receive distinct interpretations.

place of authoritative ones. Ending 2 requires more interpretative work to get (47) to come out false: Bob had to *unfix* the result of the coin flip, implying that it began fixed. So (47) is true by default. At least in this instance, we assume the default interpretation is to keep things fixed.

It may be helpful to understand the falsity of (47) after the second dialogue as an instance of the failure of antecedent strengthening—of the kind we know and love from Goodman (1947) and Lewis (1973b, 10)—but where the part that does the strengthening is implicit rather than part of the counterfactual itself.

(48) a. If Bob had bet tails, he would have lost.
b. But if Bob had bet tails [and the coin had landed tails], he would have won.

Even though the possibility of the coin landing heads is not mentioned explicitly by (47), we can formally represent its semantic role in the interpretation of that counterfactual within the second dialogue (46). We simply introduce a set IMPLICIT-HYP of implicit hypotheticals, where each 'implicit hypothetical' is a set of imaginative contents. In our illustration, let $\uparrow$ and $\downarrow$ be the imaginative contents of the coin landing heads and tails, respectively; that is, $\iota(\text{heads}) = \uparrow$ and $\iota(\text{tails}) = \downarrow$. In (45), our dialogue with the first ending, the sole implicit hypothetical is the default interpretation, to take an alternative with no imaginative content: $\textsc{implicit-hyp}_1 = \{\emptyset\}$. But in (46), our dialogue with the second ending, $\textsc{implicit-hyp}_2 = \{\{\uparrow\}, \{\downarrow\}\}$.

The two endings show how implicit hypotheticals can come to bear on our counterfactual evaluation. We take it that implicit hypotheticals achieve this by introducing counterfactual possibilities. In our modal framework this amounts to implicit hypotheticals altering the accessibility relations. We therefore reinterpret the accessibility relations as follows.

Definition 51 (Adding implicit hypotheticals)**.** *Let* IMPLICIT-HYP $\subseteq \wp(X)$ *be a nonempty set containing sets of imaginative contents. Then atomic change semantics with implicit hypotheticals is given by rewriting Definition 47 with the following clause.*

$$wR_\varphi v \quad \textit{iff} \quad v = w^{c \cup h} \textit{ for some } c \in ci(\varphi) \textit{ and } h \in \textsc{implicit-hyp}$$

We recover the original of definition of our accessibility relation under the atomic change semantics (Definition 49) by defaulting to $\textsc{implicit-hyp} = \{\emptyset\}$.

Turning back our coin flip example, Figure 5.2 depicts the resulting model in which (50) is true with respect to the implicit hypotheticals of dialogue 1 (viz. none), but false with respect to those of dialogue 2. In the diagram, the dotted arrow represents $R^1_{\text{bet T}}$, the accessibility relation of imagining at the dialogue with ending 1 if Bob had bet tails, and the dashed arrow represents $R^2_{\text{bet T}}$, the accessibility relation of imagining at ending 2 if Bob had bet tails. The straight line represents lawful temporal succession, and the actual world w is filled in.

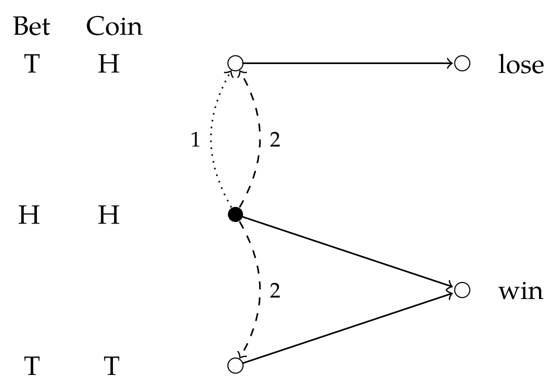

Figure 5.2: The coin flip, an example with implicit hypotheticals

The calculations behind Figure 5.2 are worked out in full in the appendix (section A9). As the diagram shows, $\{\text{bet T, coin H}\} \models \Box\text{lose}$ but $\{\text{bet T, coin T}\} \not\models \Box\text{lose}$. Then in $\mathcal{M}_1$ where $\textsc{implicit-hyp}_1 = \{\emptyset\}$, we have $\mathcal{M}_1, w \models [\text{bet T}]\Box\text{lose}$ but in $\mathcal{M}_2$ where $\textsc{implicit-hyp}_2 = \{\{\uparrow\},\{\downarrow\}\}$, we have $\mathcal{M}_2, w \not\models [\text{bet T}]\Box\text{lose}$.

The next section shows how implicit hypotheticals can invalidate a contested rule of counterfactual inference.

5.7 Case study: strong centering

This section is devoted to the following rule of counterfactual inference, which raises some intricate and instructive puzzles for the semantics of counterfactuals.

$$\frac{\varphi \qquad \psi}{[\varphi]\psi}\ (\text{g})$$

This rule of inference is validated by *strong centering*, the condition that if w already makes φ true then it is the unique world R_φ-accessible from w.

Strong centering would appear to be a characteristic property of any counterfactual semantics claiming to base itself on an intuitive notion of similarity. After all, any world differing from the actual world is surely not as similar to the actual world as the actual world is to itself. And indeed, the similarity approach of Lewis (1973b) validates (g): it is called Axiom E of a counterfactual logic that Lewis describes as, "the correct logic of counterfactual conditionals as we ordinarily understand them" (1971, 80). But the rule has some catastrophic consequences when paired with the simplification of disjunctive antecedents (a rule we encountered in section 4.2 above). Consider any sentence at all that is contingently true—say, that Sam is sitting—and any consequence of that contingency—say, that Sam's knees are bent. It is of course also true that Sam is either sitting or not sitting. But then (g) would have us infer that if Sam were sitting or not sitting, his knees would be bent, contradicting the simple observation that Sam could have been standing with straight legs. The absurdity can be pushed further in the following argument.

(49)
a. Sam is sitting or not sitting.
b. Indeed, Sam is sitting.
c. So by (g), if Sam were sitting or not sitting, he would be sitting.
d. Then by simplification, if Sam were not sitting he would be sitting.
e. And of course, if Sam were not sitting, Sam would not be sitting.
f. So if Sam were not sitting, Sam would be sitting and not sitting.

$$\dfrac{\dfrac{\dfrac{s \vee \neg s \qquad s}{[s \vee \neg s]s}\,\text{(g)}}{[\neg s]s}\,\text{Simplification} \qquad \dfrac{}{[\neg s]\neg s}\,\text{Success}}{[\neg s](s \wedge \neg s)}\,\text{Modal logic}$$

The argument's conclusion (49f) states that a contradiction would hold were Sam not sitting right now; stating, in effect, that it is impossible for Sam not to be sitting right now. Then if (g) were semantically valid, contingency would become counterfactual necessity, semantics would become metaphysics, and we would have corrupted our basic faculty of counterfactual reasoning, to imagine things other than as they happen to be—like the contingent fact of Sam sitting.

The following countermodel, depicted in Figure 5.3, shows that the atomic change semantics does not validate (g).

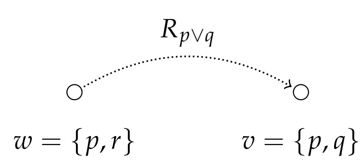

Figure 5.3: Countermodel to (g)

We let $\iota(q) = \iota(\neg r)$, and distinguish all other literals (i.e. $\iota(l) \neq \iota(l')$ for any distinct literals l, l', distinct from q and $\neg r$). By default, we put IMPLICIT-HYP $= \{\emptyset\}$. Now, according to IE's semantics $i(q) = \{\{\iota(q)\}\}$, so calculate q's changes to be $\{\iota(q)\}^+ = \{q\}$ and $\{\iota(q)\}^- = \{r\}$. We then find that $wR_q v$ as follows.

$$wR_q u \text{ iff } u = w^{\{\iota(q)\}} \text{ iff } u = (\{p, r\} - \{r\}) \cup \{q\} \text{ iff } u = \{p, q\} = v$$

Now, $i(q) \subseteq i(p) \cup i(q) = i(p \vee q)$, so $R_q \subseteq R_{p\vee q}$. Thus, $wR_{p\vee q}v$. Then $w \models p \vee q$ and $w \models r$, but $wR_{p\vee q}v$ and $v \not\models r$, i.e. $w \not\models [p \vee q]r$, contradicting (g).

The above countermodel to (g) uses a disjunctive antecedent. But one may wonder whether the problem lies not really with (g) itself but with is application to logically complex antecedents. Perhaps (g) is valid after all when φ and ψ are restricted to literal sentences. However, in his review of Lewis (1973b), Fine offers a counterexample to (g) without recourse to logically complex antecedents.

> I may speculate on a student's prospects in an exam, the results of which are already settled, and assert: if he had worked hard he would have passed. My assertion is false if the student worked hard but was only able to pass through cheating. If I say the assertion is true I cannot generalise to: if a student of similar ability had worked hard then he would have passed.
>
> (Fine, 1975, 453)

This is a compelling counterexample to (g). The student worked hard and passed, but (50) still feels false.

(50) If the student had worked hard, they would have passed.

When we judge it false that, if the student had worked hard, they would have passed, I believe we do so because when we imagine the student working hard we do not necessarily imagine them cheating. The student might have easily had a change of heart—and there is no guarantee that a student of similar ability would have cheated—so we should not take their cheating as given. This observation fits well with a key detail of Fine's scenario, that the student only passed because they cheated. In other words, if the student had worked hard but not cheated, they might have failed. Thus the counterfactual is false if, had the student worked hard, they might not have cheated—which I believe is what we have in mind in evaluating (50). After all, if the student were guaranteed to cheat then (50) would hold, as we can see from the following argument.

(51) P1 If the student had worked hard, they would have cheated.
P2 And if they had worked hard and cheated, they would have passed.
C So, if the student had worked hard, they would have passed.

Premise 1 states that if the student had worked hard, they are guaranteed to cheat. Premise 2 is justified by the claim that the student "was only able to pass through cheating" (Fine, 1975, 453), given that they worked hard. And the argument itself (51) is an application of cumulative transitivity—characterised by the axiom scheme (f)—which we took to be intuitively valid in section 4.2, and which Fine's recent counterfactual semantics also takes to be valid, given one "plausible assumption" (Fine, 2012a, 240).

Thus it is the possibility of not cheating that explains (50)'s falsity. But this possibility is not mentioned at all by (50) itself. Nonetheless, as we saw in the previous section (5.6), there are more possibilities in our intuitive counterfactual interpretation than are dreamt of by the atomic change semantics. We take it that the possibility of not cheating is an implicit hypothetical of Fine's scenario. As in the previous section, then, we introduce a set IMPLICIT-HYP containing the sets of imaginative contents we implicitly bring to bear in our counterfactual interpretation of (50).

Formally, let 'cheat', 'work', 'pass' and 'fail' be atomic sentences representing what they clearly represent. Then Figure 5.4 depicts the imaginative accessibility relation R_{work} (dotted line) and the lawful temporal succession relation (straight line).

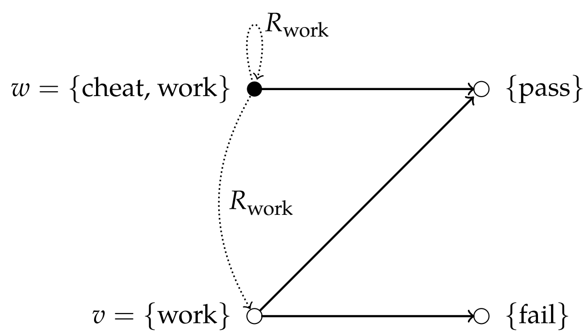

Figure 5.4: Fine's scenario with implicit hypotheticals

To see how we calculated the relations of Figure 5.4, let $\mathcal{W}$, $\mathcal{C}$ and $\mathcal{N}$ be the imaginative contents, respectively, of the student working hard, cheating and not cheating: $\iota(\text{work}) = \mathcal{W}$, $\iota(\text{cheat}) = \mathcal{C}$ and $\iota(\neg\text{cheat}) = \mathcal{N}$. Then $i(\text{work}) = \{\{\mathcal{W}\}\}$ and our implicit hypotheticals are given by IMPLICIT-HYP $= \{\{\mathcal{C}\}, \{\mathcal{N}\}\}$. (We also distinguish all other literals: $\iota(l) \neq \iota(l')$ for any distinct literals l, l' not mentioned above). We then apply Definition 47 to calculate the positive and negative changes like so.

$$\begin{aligned}
\{\mathcal{W}, \mathcal{C}\}^- &= \{p : \iota(\neg p) \in \{\mathcal{W}, \mathcal{C}\}\} && = \emptyset \\
\{\mathcal{W}, \mathcal{C}\}^+ &= \{p : \iota(p) \in \{\mathcal{W}, \mathcal{C}\}\} && = \{\text{work, cheat}\}
\end{aligned}$$

Similarly, we find $\{\mathcal{W}, \mathcal{N}\}^- = \{\text{cheat}\}$ and $\{\mathcal{W}, \mathcal{N}\}^+ = \{\text{work}\}$. Now let $w = \{\text{work, cheat}\}$ be the actual world in which the student works hard and cheats (and for simplicity, all other atomics are false). Then by our atomic change semantics with implicit hypotheticals, we calculate the accessibility relation for working hard as follows, with $w = \{\text{work, cheat}\}$.

$$
\begin{aligned}
wR_{\text{work}}u \quad & \text{iff} \quad u = w^{c \cup h} \text{ for some } c \in i(\text{work}) \text{ and } h \in \textsc{implicit-hyp} \\
& \text{iff} \quad u = w^{c \cup h} \text{ for some } c \in \{\{\mathcal{W}\}\} \text{ and } h \in \{\{\mathcal{C}\}, \{\mathcal{N}\}\} \\
& \text{iff} \quad u = w^{\{\mathcal{W}, \mathcal{C}\}} \text{ or } u = w^{\{\mathcal{W}, \mathcal{N}\}} \\
& \text{iff} \quad u = (w - \emptyset) \cup \{\text{work, cheat}\} \text{ or } u = (w - \{\text{cheat}\}) \cup \{\text{work}\} \\
& \text{iff} \quad u = \{\text{work, cheat}\} \text{ or } u = \{\text{work}\} \\
& \text{iff} \quad u = w \text{ or } u = v
\end{aligned}
$$

Hence, the two worlds that result from imagining at w that the student worked hard are w itself, and $v = \{\text{work}\}$, the world in which the student works hard but does not cheat.

Lastly, for the dynamics suppose every world where student works hard and cheats leads to passing, and that every world where they student works hard, but does not cheat, might lead to passing but might also lead to failing.

$$
\begin{aligned}
\{\text{cheat, work}\} &\rightarrow \{\text{pass}\} \\
\{\text{work}\} &\rightarrow \{\text{pass}\} \\
\{\text{work}\} &\rightarrow \{\text{fail}\}
\end{aligned}
$$

And so as desired, we find

$$\mathcal{M}, w \not\models \text{work} > \text{pass},$$

that is, $\mathcal{M}, w \not\models [\text{work}]\Box\text{pass}$, because $wR_{\text{work}}v$ and $v \rightarrow \{\text{fail}\}$. Even if the student had worked hard, they might not have cheated, and thus might not have passed, making (50) is false.

Now, the atomic change semantics leaves many questions unanswered. Some of these open questions are the subject of our next and final chapter.

Chapter 6

Puzzles

This thesis has been about the logic of imaginative content and its role in the semantic theory of counterfactuals. All in all, I hope to have accomplished three things.

1. To have axiomatised the behaviour of *and*, *or* and *not* as they appear in hypothetical contexts;
2. To have axiomatised a number of modal logics that interpret antecedents in terms of their imaginative content; and
3. To have analysed the primitives of imaginative modal logic, explaining the truth conditions of counterfactuals in terms of atomic changes to the actual world.

This chapter, on the other hand, is not about model constructions, semantic clauses or results, but to spark the reader's imagination with open questions.

6.1 Might: existential quantification over ways of imagining, over future states, or both?

Our counterfactual semantics above involved two modals, one imaginative and one temporal. According to the clause of Definition 45, a counterfactual is true just in case, for *all* ways of imagining the antecedent, and *all* paths leading from the resulting imagined world, the consequent holds at some point along the path. So we quantified over the ways of imagining the antecedent true and over the paths leading from the imagined worlds.

It is reasonable to ask what these two modalities have to say about the semantics of 'might' counterfactuals. To illustrate, suppose when Boring Bob comes to a party it is never fun, but when Fun Fatima comes it is always fun.

(52) ?? If Boring Bob or Funny Fatima had come to the party, it might have been fun.

Perhaps might-counterfactuals with disjunctive antecedents also simplify, like their would counterparts, encountered in section 4.2. If might-counterfactuals do simplify, then (52) entails the false (53), rendering (52) false by modus tollens.

(53) If Boring Bob had come to the party, it might have been fun.

However, simplicification is only valid for the universal modality $[\varphi]$, and not for the dual existential modality $\langle\varphi\rangle$: as a countermodel, consider that $\langle p \vee q\rangle q$ is valid whenever success (d) is, but we might well have $[p]\neg q$.

If might-counterfactuals with disjunctive antecedents do indeed simplify, then one could be drawn to interpreting them as $[\varphi]\Diamond\psi$, in which 'might' invokes existential quantification over future states rather than ways of imagining.

> A might-counterfactual is true just in case, for *every* way of imagining the antecedent true, there is *some* future state of the imagined world where the consequent holds.

However, some might-counterfactuals do seem to invoke existential quantification over the ways of imagining the antecedent true. Consider (54) in the context of the switches of Ciardelli et al. (2018b) (see section 1.2.2).

(54) If switch A and B were not both up, then they might both be down.

Here the might-counterfactuals appear better understood as $\langle\varphi\rangle\Diamond\psi$. We have yet to discern when to interpret a might-counterfactual as $[\varphi]\Diamond\psi$ and when as $\langle\varphi\rangle\Diamond\psi$.

6.2 Time: how long do we check for the consequent?

As we mentioned in section 4.6, the role of tense in counterfactuals is notoriously difficult. Nonetheless, in providing a semantics of counterfactuals above we have had to address some puzzles about the time at which we evaluate the consequent. Consider, for instance, (55).

(55) ?? If the lightning struck, it would strike.

Some have called (55) a tautology. Others have called it false on the grounds that lightning never strikes twice.

And here is (38), repeated as (56).

(56) ?? If it were the weekend, it would be a weekday.

On one reading, (56) is plain false: if it were the weekend it would very well be the weekend, and not a weekday. But technically, if it were the weekend it would *eventually* be a weekday. So is (56) true? Do we have habitual bounds on how far into the future we can check for the consequent? And if so, what sets the length of the bound?

One suggestion is to distinguish stative and eventive consequents.[1] The stative/eventive distinction seems to have something important to say about (57).

(57) a. EVENTIVE CONSEQUENT
If you were to flick the switch, the light would turn on.
b. STATIVE CONSEQUENT
If you were to flick the switch, the light would be on.

In light of (57), one might conjecture that stative consequents must hold *immediately* at the moment of change for the counterfactual to come out true, whereas eventive consequents must only hold at *some* point in the future for the counterfactual to hold. But focusing on the distinction between eventive and stative consequents is just one suggestion.

6.3 Counterfactuals as entailments from imperatives to imagine?

Following the work of Clark-Younger (2014) and Parsons (2013), it is reasonable to assume that imperatives can feature in entailment relations, as in (58).

(58) Have a nap! Then you will feel better.

As well as imperatives to action we can also consider imperatives to *imagine*; (59), for example.

(59) Imagine you had a nap!

It seems imperatives to imagine can feature in entailment relations just like any other kind of imperative—but with the consequent taking on an X-marking, as in (60), with 'would' in place of 'will' indicating that we evaluate the consequent with respect to a hypothetical context.[2]

(60) Imagine you had a nap. Then you would feel better.

Now compare the semantic contribution of (60) with (61).

[1] Thanks to Morwenna Hoeks and Jonathan Pesetsky for this suggestion.
[2] The term 'X-marking' comes from Sabine Iatridou (see, e.g., Iatridou and von Fintel, 2017).

(61) If you had a nap, then you would feel better.

There appears to be a tantalisingly close relationship between counterfactuals and entailments from imperatives to imagine. The imperative character of counterfactual antecedents offers hope in explaining why counterfactual antecedents behave in the way they do. For example, we saw in section 2.5.1 that disjunction in hypothetical contexts is peculiarly neither inclusive nor exclusive. Analogously, the imperatives in (62) intuitively all have distinct meanings.

(62) a. Flick switch A down or B down! PLAIN
b. Flick switch A down, or B down, or both down! INCLUSIVE
c. Flick switch A down and B up, or A up and B down! EXCLUSIVE

Though without a general semantics of imperatives to hand, this is not the place to argue about the exact relationship between counterfactuals on the one hand, and entailments from imperatives to imagine on the other. But recent work on the semantics of imperatives (e.g. Aloni, 2007; Aloni and Ciardelli, 2011; von Fintel and Iatridou, 2017) opens up the avenue for a comparison of their respective semantic contributions—which might even turn out to be identical. A semantics of counterfactual antecedents via imperatives may well elucidate some of the quirks of counterfactuals in ways we have yet to imagine.

Appendix

A1 De Morgan's first law in truthmaker semantics

A model according to truthmaker semantics (Fine, 2016, 205) is defined to be a triple $(S, \sqsubseteq, [\![\cdot]\!])$, where

- S is a nonempty set.
- $\sqsubseteq$ is a partial order (reflexive, transitive, antisymmetric) such that every nonempty subset of S with a bound has a least upper bound.
- $[\![\cdot]\!] : At \to \wp(S) \times \wp(S)$ is a valuation assigning to each atomic sentence p a pair (V, F), for $V, F \subseteq S$.

For a discussion of what states and fusions are, see (Fine, 2012a, 233–236). In proving the fact that De Morgan's first law is valid with respect to truthmaker semantics, we will not need to go into any specifics of what Fine takes fusions and states to be—the result will follow from the semantic clauses for the connectives alone.

Fine (2012a, 234) provides the following clauses for negation, conjunction and disjunction, according to his *inclusive* truthmaker semantics. In the clauses below, we take "verifies" to mean "exactly verifies"; i.e. "is a truthmaker for".

(i)$^+$ a state s verifies $\neg\varphi$ if and only if (iff) s falsifies φ;
(i)$^-$ s falsifies $\neg\varphi$ iff s verifies φ;
(ii)$^+$ s verifies $\varphi \wedge \psi$ iff s is the fusion $s_1 \sqcup s_2$ of a state s_1 that verifies φ and a state s_2 that verifies ψ;
(ii)$^-$ s falsifies $\varphi \wedge \psi$ iff s falsifies φ or s falsifies ψ or s falsifies $\varphi \vee \psi$;
(iii)$^+$ s verifies $\varphi \vee \psi$ iff s verifies φ or s verifies ψ or s verifies $\varphi \wedge \psi$;
(iii)$^-$ s falsifies $\varphi \vee \psi$ iff s is the fusion $s_1 \sqcup s_2$ of a state s_1 that falsifies φ and a state s_2 that falsifies ψ.

Fine (2012a, 235) also suggests considering a more exclusive version of the clauses above. We achieve exclusive truthmaker semantics by replacing clauses (ii)$^-$ and (iii)$^+$ above, respectively, with the following:

(ii)$^-_{\text{ex}}$ s falsifies $\varphi \wedge \psi$ iff s falsifies φ or s falsifies ψ;
(iii)$^+_{\text{ex}}$ s verifies $\varphi \vee \psi$ iff s verifies φ or s verifies ψ.

In addition to exact verification, Fine has also considered two notions of *inexact* and *loose* verification (see Fine, 2014), each weaker than the last (i.e. for any sentence φ, every exact verifier for φ is also an inexact verifier for φ, and every inexact verifier for φ is also a loose verifier for φ). Thus, to prove that De Morgan's first law is valid with respect to all three notions of verification, it suffices to show that it is valid with respect to exact verification. Let us use $s \Vdash \varphi$ to indicate that a state s exactly verifies a sentence φ, and $s \dashv\!| \varphi$ to indicate that s exactly falsifies φ.

Inclusive truthmaker semantics

Fact 52 (De Morgan's first law, inclusively). *In any model and state s of inclusive truthmaker semantics, s verifies $\neg(\varphi \wedge \psi)$ if and only if it verifies $\neg\varphi \vee \neg\psi$.*

Proof. Observe the following chain of equivalences.

$$
\begin{array}{lll}
 & s \Vdash \neg(\varphi \wedge \psi) & \\
\text{iff} & s \dashv\!| \varphi \wedge \psi & \text{(i)}^- \\
\text{iff} & s \dashv\!| \varphi \text{ or } s \dashv\!| \psi \text{ or } s \dashv\!| \varphi \vee \psi & \text{(ii)}^- \\
\text{iff} & s \dashv\!| \varphi \text{ or } s \dashv\!| \psi \text{ or } s = s_1 \sqcup s_2 \text{ where } s_1 \dashv\!| \varphi \text{ and } s_2 \dashv\!| \psi & \text{(iii)}^- \\
\text{iff} & s \Vdash \neg\varphi \text{ or } s \Vdash \neg\psi \text{ or } s = s_1 \sqcup s_2 \text{ where } s_1 \Vdash \neg\varphi \text{ and } s_2 \Vdash \neg\psi & \text{(i)}^+ \\
\text{iff} & s \Vdash \neg\varphi \text{ or } s \Vdash \neg\psi \text{ or } s \Vdash \neg\varphi \wedge \neg\psi & \text{(ii)}^+ \\
\text{iff} & s \Vdash \neg\varphi \vee \neg\psi & \text{(iii)}^+
\end{array}
$$

□

Exclusive truthmaker semantics It is even easier to show that $\neg(\varphi \wedge \psi)$ and $\neg\varphi \vee \neg\psi$ are equivalent with respect to the exclusive clauses above.

Fact 53 (De Morgan's first law, exclusively). *In any model and state s of exclusive truthmaker semantics, s verifies $\neg(\varphi \wedge \psi)$ if and only if it verifies $\neg\varphi \vee \neg\psi$.*

Proof. Observe the following chain of equivalences.

$$
\begin{array}{lll}
 & s \Vdash \neg(\varphi \wedge \psi) & \\
\text{iff} & s \dashv\!| \varphi \wedge \psi & \text{(i)}^- \\
\text{iff} & s \dashv\!| \varphi \text{ or } s \dashv\!| \psi & \text{(ii)}^-_{\text{ex}} \\
\text{iff} & s \Vdash \neg\varphi \text{ or } s \Vdash \neg\psi & \text{(i)}^+ \\
\text{iff} & s \Vdash \neg\varphi \vee \neg\psi & \text{(iii)}^+_{\text{ex}}
\end{array}
$$

□

A2 Substitution in IE

Let us say that ψ *occurs under the scope of a negation in* φ just in case ψ is a subformula of $\neg\chi$ for some subformula $\neg\chi$ of φ.

Fact 54. *Unnegated equivalents are salva aequivalente substitutable. That is,*

$$\frac{\psi \equiv \chi}{\varphi \equiv \varphi[\chi/\psi]}$$

is a derived rule of IE, where ψ does not occur under the scope of a negation in φ.

Proof. The proof relies essentially on $\vee$ and $\wedge$ addition. Assume $\psi \equiv \chi$ as a premise. We distinguish two cases: either φ is ψ or it is not. If $\varphi = \psi$ then $\varphi[\chi/\psi] = \chi$, and so $\varphi = \psi \equiv \chi = \varphi[\chi/\psi]$, as required. So suppose $\varphi \neq \psi$, and that ψ does not occur under the scope of $\neg$ in φ. We show $\varphi \equiv \varphi[\chi/\psi]$ by induction on the complexity of φ.

If $\varphi = p$ then as $p \neq \psi$, $p[\chi/\psi] = p \equiv p$ by Fact 1. If φ is $\neg\sigma$ then as ψ does not occur under the scope of $\neg$ in φ, ψ does not appear in φ at all. Hence $\varphi \equiv \varphi[\chi/\psi]$.

If $\varphi = \sigma \vee \tau$ then as $\varphi \neq \psi$, $\psi \neq (\sigma \vee \tau)$ and so $(\sigma \vee \tau)[\chi/\psi] = \sigma[\chi/\psi] \vee \tau[\chi/\psi]$. As ψ does not appear under $\neg$ in φ, it also does not appear under $\neg$ in either σ or τ. Then by induction hypothesis, $\sigma \equiv \sigma[\chi/\psi]$ and $\tau \equiv \sigma[\chi/\psi]$. We quickly verify that

$$\frac{\gamma \vee \gamma' \equiv \delta \vee \delta'}{\gamma' \vee \gamma \equiv \delta' \vee \delta}$$

is a derived rule, for any sentences $\gamma, \gamma', \delta, \delta'$.

$$\dfrac{\dfrac{\dfrac{}{\gamma' \vee \gamma \equiv \gamma \vee \gamma'}\text{A4} \qquad \gamma \vee \gamma' \equiv \delta \vee \delta'}{\gamma' \vee \gamma \equiv \delta \vee \delta'}\text{Trans.} \qquad \dfrac{}{\delta \vee \delta' \equiv \delta' \vee \delta}\text{A4}}{\gamma' \vee \gamma \equiv \delta' \vee \delta}\text{Trans.}$$

We then derive the result using $\vee$ addition, as follows.

$$\dfrac{\dfrac{\dfrac{\psi \equiv \chi}{\sigma \equiv \sigma[\chi/\psi]}\text{I.H.}}{\sigma \vee \tau \equiv \sigma[\chi/\psi] \vee \tau}\vee\text{ addition} \qquad \dfrac{\dfrac{\dfrac{\psi \equiv \chi}{\tau \equiv \tau[\chi/\psi]}\text{I.H.}}{\tau \vee \sigma[\chi/\psi] \equiv \tau[\chi/\psi] \vee \sigma[\chi/\psi]}\vee\text{ addition}}{\sigma[\chi/\psi] \vee \tau \equiv \sigma[\chi/\psi] \vee \tau[\chi/\psi]}\text{Above}}{\sigma \vee \tau \equiv \sigma[\chi/\psi] \vee \tau[\chi/\psi]}\text{Transitivity}$$

Hence $\varphi \equiv \varphi[\chi/\psi]$ as $\varphi = \sigma \vee \tau \equiv \sigma[\chi/\psi] \vee \tau[\chi/\psi] = (\sigma \vee \tau)[\chi/\psi] = \varphi[\chi/\psi]$.

The case where φ is a conjunction is similar, this time using communitativity of $\wedge$ (A5) to show $\gamma \wedge \gamma' \equiv \delta \wedge \delta' \vdash \gamma' \wedge \gamma \equiv \delta' \wedge \delta$ and then using $\wedge$ addition. $\square$

A3 Soundness of IE

To prove soundness we make use of the following four facts.

Fact 55 (Distribution of $\wedge$ over $\vee$)**.** $i\big((\varphi \wedge (\psi \vee \chi)\big) = i\big((\varphi \wedge \psi) \vee (\varphi \wedge \chi)\big)$.

Proof. The result from distribution of classical 'and' over 'or'. To be explicit, we unpack the clauses for disjunction and conjuncton as follows.

$$\begin{aligned}
i(\varphi \wedge (\psi \vee \chi)) &= i(\varphi) \Cup (i(\psi) \cup i(\chi)) \\
&= \{a \cup b : a \in i(\varphi) \text{ and } (b \in i(\psi) \text{ or } b \in i(\chi))\} \\
&= \{a \cup b : (a \in i(\varphi), b \in i(\psi)) \text{ or } (a \in i(\varphi), b \in i(\chi))\} \\
&= \{a \cup b : a \in i(\varphi), b \in i(\psi))\} \cup \{a \cup b : a \in i(\varphi), b \in i(\chi))\} \\
&= (i(\varphi) \Cup i(\psi)) \cup (i(\varphi) \Cup i(\chi)) \\
&= i((\varphi \wedge \psi) \vee (\varphi \wedge \chi))
\end{aligned}$$

□

In the following three facts, note that by the semantic clauses of IE, $i(\varphi)$ is a nonempty set of nonempty sets, for every model of IE and sentence $\varphi \in \mathcal{L}$.

Fact 56 (Idempotence with respect to $\cup$ and $\Cup$)**.** *For any set A recall that $A^{\cup} = \{\bigcup A' : A' \subseteq A, A' \neq \emptyset\}$ is A's closure under union. Then for any sets A and B, closure under union is idempotent with respect to union and pairwise union, in the following sense.*

a. $(A^{\cup} \cup B)^{\cup} = (A \cup B)^{\cup}$
b. $(A^{\cup} \Cup B)^{\cup} = (A \Cup B)^{\cup}$

Hence

c. $(A^{\cup} \cup B^{\cup})^{\cup} = (A \cup B)^{\cup}$
d. $(A^{\cup} \Cup B^{\cup})^{\cup} = (A \Cup B)^{\cup}$

Proof. For (a), we have to show that

$$\{\bigcup C : C \subseteq A^{\cup} \cup B, C \neq \emptyset\} = \{\bigcup D : D \subseteq A \cup B, D \neq \emptyset\}.$$

The left-to-right inclusion amounts to showing that for every nonempty $C \subseteq A^{\cup} \cup B$ there is a nonempty $D \subseteq A \cup B$ such that $\bigcup C = \bigcup D$. For arbitrary such C, take $D = \{d \in A \cup B : d \subseteq \bigcup C\}$. Then clearly $D \subseteq A \cup B$, D is nonempty, and $\bigcup C = \bigcup D$. Conversely, the right-to-left inclusion amounts to showing that for every nonempty $D \subseteq A \cup B$ there is a nonempty $C \subseteq A^{\cup} \cup B$ such that $\bigcup C = \bigcup D$. Now, as $A \subseteq A^{\cup}$ we have that $D \subseteq A \cup B$ implies $D \subseteq A^{\cup} \cup B$. So simply take $C = D$.

Part (b) is similar. For the left-to-right inclusion, given any nonempty $C \subseteq A^{\cup} \Cup B$ take $D = \{d \in A \Cup B : d \subseteq \bigcup C\}$. Then as C is nonempty, A and B are nonempty too (since $x \Cup \emptyset = \emptyset$ for every set x). Then D is nonempty and, by construction, $D \subseteq A \Cup B$ and $\bigcup C = \bigcup D$. For the right-to-left inclusion, note again that as $A \subseteq A^{\cup}$, we have that $D \subseteq A \Cup B$ implies $D \subseteq A^{\cup} \Cup B$, so simply take $C = D$.

Applying (a) twice, we see $(A^{\cup} \cup B^{\cup})^{\cup} = (A \cup B^{\cup})^{\cup} = (A \cup B)^{\cup}$. Hence (c). We similarly apply (b) twice to derive (d). $\square$

Fact 57. *$(A \cup B)^{\cup} = A^{\cup} \cup B^{\cup} \cup (A^{\cup} \Cup B^{\cup})$ for any sets A and B.*

Proof. First recall that $(A \cup B)^{\cup} = \{\bigcup C : C \subseteq A \cup B, C \neq \emptyset\}$.

($\subseteq$) We show that $\{\bigcup C : C \subseteq A \cup B\} \subseteq A^{\cup} \cup B^{\cup} \cup (A^{\cup} \Cup B^{\cup})$. Take any nonempty $C \subseteq A \cup B$. Then either C is disjoint from A, from B, or from neither. Since $C \subseteq A \cup B$, if $A \cap C = \emptyset$ then $C \subseteq B$, and so $\bigcup C \in B^{\cup}$, as desired. Likewise, if $B \cap C = \emptyset$ then $\bigcup C \in A^{\cup}$, as desired.

So suppose C is disjoint from neither A nor B; that is, $A \cap C \neq \emptyset \neq B \cap C$. Then A and B are nonempty, and so $A^{\cup}$ and $B^{\cup}$ are nonempty too. Thus, $A^{\cup} \Cup B^{\cup}$ is nonempty. We show that $\bigcup C \in (A^{\cup} \Cup B^{\cup})$. Note,

$$\begin{aligned}(A^{\cup} \Cup B^{\cup}) &= \{a \cup b : a \in A^{\cup}, b \in B^{\cup}\} \\ &= \{\bigcup A' \cup \bigcup B' : A' \subseteq A, B' \subseteq B, A' \neq \emptyset \neq B'\}.\end{aligned}$$

Of course, $A \cap C \subseteq A$ and $B \cap C \subseteq B$, so take $A' = A \cap C$ and $B' = B \cap C$. Then as $C \subseteq A \cup B$, $C = (A \cap C) \cup (B \cap C)$, so we have

$$\bigcup C = \bigcup((A \cap C) \cup (B \cap C)) = \bigcup(A \cap C) \cup \bigcup(B \cap C).$$

Hence $\bigcup C \in (A^{\cup} \Cup B^{\cup})$, as desired.

($\supseteq$) To show that $A^{\cup} \subseteq (A \cup B)^{\cup}$, observe the following chain of inclusions, $A^{\cup} \subseteq A^{\cup} \cup B \subseteq (A^{\cup} \cup B)^{\cup} = (A \cup B)^{\cup}$, with Fact 56(a) giving the final identity. Similarly, $B^{\cup} \subseteq (A \cup B)^{\cup}$, this time using Fact 56(b).

It remains to show $A^{\cup} \Cup B^{\cup} \subseteq (A \cup B)^{\cup}$. As $A^{\cup} \Cup B^{\cup} = \{\bigcup A' \cup \bigcup B' : A' \subseteq A, B' \subseteq B, A' \neq \emptyset \neq B'\}$, this amounts to showing that for every nonempty $A' \subseteq A$ and nonempty $B' \subseteq B$ there is a nonempty $C \subseteq A \cup B$ such that $\bigcup A' \cup \bigcup B' = \bigcup C$. Taking $C = A' \cup B'$ will do: since $A' \subseteq A$ and $B' \subseteq B$, $A' \cup B' \subseteq A \cup B$, as required. And clearly, $\bigcup(A' \cup B') = \bigcup A' \cup \bigcup B'$.

$\square$

Fact 58 (Closure under union distributes over $\Cup$)**.** $(A \Cup B)^{\cup} = (A^{\cup} \Cup B^{\cup})$

Proof. First unpacking the definitions, observe that

$$\begin{aligned} A^{\cup} \Cup B^{\cup} &= \{\textstyle\bigcup A' : A' \subseteq A, A' \neq \emptyset\} \Cup \{\textstyle\bigcup B' : B' \subseteq B, B' \neq \emptyset\} \\ &= \{\textstyle\bigcup A' \cup \bigcup B' : A' \subseteq A, B' \subseteq B, A' \neq \emptyset \neq B'\} \\ &= \{\textstyle\bigcup (A' \cup B') : A' \subseteq A, B' \subseteq B, A' \neq \emptyset \neq B'\}. \end{aligned}$$

($\subseteq$) We have to show that $\{\bigcup C : C \subseteq A \Cup B\} \subseteq \{\bigcup(A' \cup B') : A' \subseteq A, B' \subseteq B\}$. That is, for every nonempty $C \subseteq A \Cup B$ there are nonempty $A' \subseteq A$ and $B' \subseteq B$ such that $\bigcup C = \bigcup(A' \cup B')$. So pick any $C \subseteq A \Cup B$. It will suffice to take $A' = \{a \in A : a \subseteq \bigcup C\}$ and $B' = \{a \in B : b \subseteq \bigcup C\}$. Then clearly, A' and B' are nonempty, $A' \subseteq A$ and $B' \subseteq B$. It remains to show that $\bigcup C = \bigcup(A' \cup B')$. It is easy to check that, as $C \subseteq A \Cup B$, $\bigcup C \subseteq \bigcup A'$ and $\bigcup C \subseteq \bigcup B'$ by construction of A' and B'. Thus $\bigcup C \subseteq \bigcup A' \cup \bigcup B' = \bigcup(A' \cup B')$. Conversely, for any $x \in \bigcup(A' \cup B')$, $x \in y \subseteq \bigcup C$ for some $y \in A \cup B$ by construction of A' and B', so $x \in \bigcup C$. Hence $\bigcup C = \bigcup(A' \cup B')$, as required.

($\supseteq$) Simply observe that $A^{\cup} \Cup B^{\cup} \subseteq (A^{\cup} \Cup B^{\cup})^{\cup} = (A \Cup B)^{\cup}$ by Fact 56(b).

□

Theorem 59 (Soundness). *Every theorem of IE is valid.*

Proof. We show that the axioms are valid and that the rules of inference perserve validity. Pick any model $(X, \imath)$ of IE.

(A1) $i(\neg\neg p) = i^{-}(\neg p)^{\cup} = i(p)^{\cup}$, but as $i(p)$ is a singleton, it is already closed under union. Thus, $i(p)^{\cup} = i(p)$.

(A2) As $i(p)$ is a singleton, $i(p \wedge p) = i(p) \Cup i(p) = i(p)$.

(A3–8) Immediate from properties of union, with (A8) given by Fact 55.

(A9) $i(\neg(\varphi \wedge \psi)) = i^{-}(\varphi \wedge \psi)^{\cup} = \big(i^{-}(\varphi) \cup i^{-}(\psi)\big)^{\cup}$ which by Fact 57 is

$$\begin{aligned} &= i^{-}(\varphi)^{\cup} \cup i^{-}(\psi)^{\cup} \cup \big(i^{-}(\varphi)^{\cup} \Cup i^{-}(\psi)^{\cup}\big) \\ &= \big(i(\neg\varphi) \cup i(\neg\psi)\big) \cup \big(i(\neg\varphi) \Cup i(\neg\psi)\big) \\ &= i(\neg\varphi \vee \neg\psi) \cup i(\neg\varphi \wedge \neg\psi) \\ &= i\big((\neg\varphi \vee \neg\psi) \vee (\neg\varphi \wedge \neg\psi)\big). \end{aligned}$$

(A10) $i(\neg(\varphi \vee \psi)) = i^{-}(\varphi \vee \psi)^{\cup} = \big(i^{-}(\varphi) \Cup i^{-}(\psi)\big)^{\cup}$ which by Fact 58 is $i^{-}(\varphi)^{\cup} \Cup i^{-}(\psi)^{\cup} = i(\neg\varphi) \Cup i(\neg\psi) = i(\neg\varphi \wedge \neg\psi)$.

(A11) Follows from the idempotence of closure under union; i.e. $A^{\cup\cup} = A^{\cup}$. $i(\neg\neg\neg\varphi) = i^{-}(\neg\neg\varphi)^{\cup} = i(\neg\varphi)^{\cup} = i^{-}(\varphi)^{\cup\cup} = i^{-}(\varphi)^{\cup} = i(\neg\varphi)$.

(A12) $i(\neg\neg(\varphi \wedge \psi)) = i^-(\neg(\varphi \wedge \psi))^\cup = i(\varphi \wedge \psi)^\cup = \big(i(\varphi) \Cup i(\psi)\big)^\cup$ which by Fact 58 is $i(\varphi)^\cup \Cup i(\psi)^\cup = i^-(\neg\varphi)^\cup \Cup i^-(\neg\psi)^\cup = i(\neg\neg\varphi) \Cup i(\neg\neg\psi) = i(\neg\neg\varphi \wedge \neg\neg\psi)$.

(A13) $i\big(\neg\neg(\varphi \vee \psi)\big) = i^-\big(\neg(\varphi \vee \psi)\big)^\cup = i(\varphi \vee \psi)^\cup = \big(i(\varphi) \cup i(\psi)\big)^\cup = \big(i^-(\neg\varphi) \cup i^-(\neg\psi)\big)^\cup = i^-(\neg\varphi \wedge \neg\psi)^\cup = i\big(\neg(\neg\varphi \wedge \neg\psi)\big)$.

(Rules) Symmetry and Transitivity follow respectively from the symmetry and transitivity of identity. The addition rules are immediate.

□

A4 Facts about imaginative inclusion

Fact 60. *The following are derived rules of IE + (Definition of $\hookrightarrow$),*

$$\frac{\varphi \hookrightarrow \psi \qquad \psi \hookrightarrow \varphi}{\varphi \equiv \psi}\ \equiv \textit{ introduction} \qquad\qquad \frac{\varphi \equiv \psi}{\varphi \hookrightarrow \psi}\ \equiv \textit{ elimination}$$

$$\frac{\varphi \hookrightarrow \psi \qquad \psi \hookrightarrow \chi}{\varphi \hookrightarrow \chi}\ \textit{Transitivity of} \hookrightarrow \qquad \frac{\varphi \vee \psi \hookrightarrow \chi}{\varphi \hookrightarrow \chi}\ \textit{Simplification}$$

and $\varphi \hookrightarrow \varphi \vee \psi$ is a theorem of IE + (Definition of $\hookrightarrow$).

Proof. From $\varphi \vee \psi \equiv \psi$, using (A4), symmetry and transitivity one can easily derive $\psi \equiv \psi \vee \varphi$.

$$\dfrac{\dfrac{\dfrac{\dfrac{\varphi \hookrightarrow \psi}{\varphi \vee \psi \equiv \psi}\ \text{Def.} \hookrightarrow}{\psi \equiv \psi \vee \varphi}\ \text{A4, Sym., Trans.} \qquad \dfrac{\psi \hookrightarrow \varphi}{\psi \vee \varphi \equiv \varphi}\ \text{Def.} \hookrightarrow}{\psi \equiv \varphi}\ \text{Transitivity}}{\varphi \equiv \psi}\ \text{Symmetry}$$

To prove transitivity of $\hookrightarrow$, we apply $\vee$ addition as follows.

$$\dfrac{\dfrac{\dfrac{\dfrac{\psi \hookrightarrow \chi}{\chi \equiv \psi \vee \chi}\ \text{Def.} \hookrightarrow\text{, sym.}}{\varphi \vee \chi \equiv \varphi \vee \psi \vee \chi}\ \vee\text{ add.} \qquad \dfrac{\dfrac{\dfrac{\varphi \hookrightarrow \psi}{\varphi \vee \psi \equiv \psi}\ \text{Def.} \hookrightarrow}{\varphi \vee \psi \vee \chi \equiv \psi \vee \chi}\ \vee\text{ add.} \qquad \dfrac{\psi \hookrightarrow \chi}{\psi \vee \chi \equiv \chi}\ \text{Def.} \hookrightarrow}{\varphi \vee \psi \vee \chi \equiv \chi}\ \text{Trans.}}{\varphi \vee \chi \equiv \chi}\ \text{Trans.}}{\varphi \hookrightarrow \chi}\ \text{Def.} \hookrightarrow$$

We next show that imaginative inclusion follows from imaginative entailment.

$$\cfrac{\cfrac{\cfrac{\cfrac{}{\varphi \equiv \varphi \vee \varphi}\,\text{A3} \qquad \cfrac{\varphi \equiv \psi}{\varphi \vee \varphi \equiv \psi \vee \varphi}\,\vee\text{ addition}}{\varphi \equiv \psi \vee \varphi}\,\text{Transitivity}}{\varphi \equiv \varphi \vee \psi}\,\text{Symmetry}}{\varphi \hookrightarrow \psi}\,\text{Def. }\hookrightarrow$$

The proof that $\varphi \hookrightarrow \varphi \vee \psi$ is a theorem also uses $\vee$ addition.

$$\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{}{\varphi \equiv \varphi \vee \varphi}\,\text{A3}}{\varphi \vee \varphi \equiv \varphi}\,\text{Symmetry}}{(\varphi \vee \varphi) \vee \psi \equiv \varphi \vee \psi}\,\vee\text{ addition}}{\varphi \vee (\varphi \vee \psi) \equiv \varphi \vee \psi}\,\text{A6}}{\varphi \hookrightarrow \varphi \vee \psi}\,\text{Definition of }\hookrightarrow$$

From this and transitivity, the remaining rule follows at once.

$$\cfrac{\cfrac{}{\varphi \hookrightarrow \varphi \vee \psi}\,\text{Above} \qquad \varphi \vee \psi \hookrightarrow \chi}{\varphi \hookrightarrow \chi}\,\hookrightarrow\text{ transitivity}$$

□

A5 Soundness of CIE

Fact 61 (Inconsistency is persistent over conjunction)**.** *For any $\varphi, \psi \in \mathcal{L}$ and any model (X, ι) of IE, if $i(\varphi)$ is inconsistent w.r.t. (X, ι), then so is $i(\varphi \wedge \psi)$.*

Proof. If $i(\varphi)$ is empty then so is $i(\varphi) \Cup i(\psi) = i(\varphi \wedge \psi)$, in which case $i(\varphi \wedge \psi)$ is vacuously inconsistent. So suppose $i(\varphi)$ is nonempty and pick any $x \in i(\varphi \wedge \psi)$. Then $x = a \cup b$ for some $a \in i(\varphi)$ and $b \in i(\psi)$. Then as $i(\varphi)$ is inconsistent, a is too, and as $a \subseteq x$, so is x. Since x was arbitrary, $i(\varphi \wedge \psi)$ is inconsistent. □

Lemma 62. *For any model (X, ι) of IE and $A, B \subseteq \wp(X)$, if $A \Cup B$ is inconsistent with respect to (X, ι) then so is $A^{\cup} \Cup B$.*

Proof. If $A \Cup B$ is empty then so is $A^{\cup} \Cup B$, in which case $A^{\cup} \Cup B$ is vacuously inconsistent. So suppose $A \Cup B$ is nonempty and pick any $x \in A^{\cup} \Cup B$. Then clearly $x \supseteq y$ for some $y \in A \Cup B$. Then as $A \Cup B$ is inconsistent, y is too, and as $y \subseteq x$, x is also inconsistent. Since x was arbitrary, $A^{\cup} \Cup B$ is inconsistent. □

Fact 63 (Syntactic contradictions are semantic inconsistencies)**.** *For every φ of $\mathcal{L}$ and model (X, ι) of IE, $i(\varphi \wedge \neg\varphi)$ is inconsistent with respect to (X, ι).*

Proof. By induction on the complexity of φ.

(φ is p). Then $i(p \wedge \neg p) = \{\{\iota(p), \iota(\neg p)\}\}$, which is inconsistent since, by definition, $\{\iota(p), \iota(\neg p)\}$ is inconsistent.

(φ is $\neg\psi$). We have to show that $i(\psi)^{\cup} \Cup i^{-}(\psi)^{\cup}$ is inconsistent, since

$$i(\neg\psi \wedge \neg\neg\psi) = i(\neg\psi) \Cup i(\neg\neg\psi) = i^{-}(\psi)^{\cup} \Cup i^{-}(\neg\psi)^{\cup} = i^{-}(\psi)^{\cup} \Cup i(\psi)^{\cup}.$$

By induction hypothesis, $i(\psi \wedge \neg\psi) = i(\psi) \Cup i^{-}(\psi)^{\cup}$ is inconsistent. Then by the previous Lemma (62), $i(\psi)^{\cup} \Cup i^{-}(\psi)^{\cup}$ is inconsistent, as required.

(φ is $\psi \wedge \chi$). First observe the following chain of equivalences of IE.

$$\begin{aligned}
\varphi \wedge \neg\varphi &\equiv (\psi \wedge \chi) \wedge \neg(\psi \wedge \chi) \\
&\equiv (\psi \wedge \chi) \wedge (\neg\psi \vee \neg\chi \vee (\neg\psi \wedge \neg\chi)) && \text{(A9)} \\
&\equiv (\psi \wedge \chi \wedge \neg\psi) \vee (\psi \wedge \chi \wedge \neg\chi) \vee (\psi \wedge \chi \wedge \neg\psi \wedge \neg\chi) && \text{(A8)}
\end{aligned}$$

So by soundness of IE (Theorem 59), $i(\varphi \wedge \neg\varphi) = i\big((\psi \wedge \chi \wedge \neg\psi) \vee (\psi \wedge \chi \wedge \neg\chi) \vee (\psi \wedge \chi \wedge \neg\psi \wedge \neg\chi)\big)$. Now by induction hypothesis, $i(\psi \wedge \neg\psi)$ and $i(\chi \wedge \neg\chi)$ are inconsistent. Then by Fact 61, the three disjuncts $(\psi \wedge \chi \wedge \neg\psi)$, $(\psi \wedge \chi \wedge \neg\chi)$ and $(\psi \wedge \chi \wedge \neg\psi \wedge \neg\chi)$ are all inconsistent. Then any $x \in i(\varphi \wedge \neg\varphi)$ is an element of at least one of the disjuncts, and so is inconsistent too. Hence $i(\varphi \wedge \neg\varphi)$ is inconsistent.

(φ is $\psi \vee \chi$). Observe the following equivalences of IE.

$$\begin{aligned}
\varphi \wedge \neg\varphi &\equiv (\psi \vee \chi) \wedge \neg(\psi \vee \chi) \\
&\equiv (\psi \vee \chi) \wedge (\neg\psi \wedge \neg\chi) && \text{(A10)} \\
&\equiv (\psi \wedge \neg\psi \wedge \neg\chi) \vee (\chi \wedge \neg\psi \wedge \neg\chi) && \text{(A8)}
\end{aligned}$$

So soundness of IE, $i(\varphi \wedge \neg\varphi) = i\big((\psi \wedge \neg\psi \wedge \neg\chi) \vee (\chi \wedge \neg\psi \wedge \neg\chi)\big)$. Now induction hypothesis, $i(\psi \wedge \neg\psi)$ and $i(\chi \wedge \neg\chi)$ are inconsistent. Then by Fact 61, so are $i(\psi \wedge \neg\psi \wedge \neg\chi)$ and $i(\chi \wedge \neg\psi \wedge \neg\chi)$. Then any $x \in i(\varphi \wedge \neg\varphi)$ is inconsistent, and so $i(\varphi \wedge \neg\varphi)$ is itself inconsistent.

$\square$

A6 Soundness of IPL

Recall IPL's additional three rules.

$$\frac{\alpha \hookrightarrow \gamma}{\gamma \hookrightarrow \alpha}\ \text{Literal symmetry}$$

$$\frac{\gamma \hookrightarrow (\alpha_1 \vee \cdots \vee \alpha_n)}{(\gamma \hookrightarrow \alpha_1) \vee \cdots \vee (\gamma \hookrightarrow \alpha_n)}\ \vee\ \text{distribution}$$

$$\frac{(\alpha \wedge l) \hookrightarrow (l_1 \wedge \cdots \wedge l_n)}{(l \hookrightarrow l_1) \vee \cdots \vee (l \hookrightarrow l_n)}\ \wedge\ \text{distribution}$$

Theorem 64 (Soundness of IPL)**.** *Every theorem of IPL is true in every model of IPL.*

Proof. Pick any model $w = (\mathcal{I}, X, \iota)$ of IPL. Let $\alpha, \alpha_1, \ldots, \alpha_n, \gamma, l, l_1, \ldots, l_n \in \mathcal{L}$ where $l, l_1, \ldots, l_n$ are literals and γ is a conjunction of literals. (Note we may restrict attention to $\mathcal{L}$ since each sentence occurs under the scope of $\hookrightarrow$ in the rules above.) By the semantic clauses of IE, $i(\alpha)$ is nonempty, and both $i(\gamma)$ and $i(l_1 \wedge \cdots \wedge l_n)$ are singletons.

Literal symmetry. Suppose $w \models \alpha \hookrightarrow \gamma$, i.e. $i(\alpha) \subseteq i(\gamma)$. Then as $i(\gamma)$ is a singleton and $i(\alpha)$ nonempty, $i(\alpha) = i(\gamma)$; in particular, $i(\gamma) \subseteq i(\alpha)$, so $w \models \gamma \hookrightarrow \alpha$.

$\vee$ *distribution*. Suppose $i(\gamma) \subseteq i(\alpha_1 \vee \cdots \vee \alpha_n)$. As $i(\gamma)$ is a singleton, say $i(\gamma) = \{Y\}$ for some $Y \subseteq X$. Then $Y \in i(\alpha_k)$ for some $k \leq n$, and so $w \models \gamma \hookrightarrow \alpha_k$. Hence $w \models (\gamma \hookrightarrow \alpha_1) \vee \cdots \vee (\gamma \hookrightarrow \alpha_n)$.

$\wedge$ *distribution*. Suppose $i(\alpha \wedge l) \subseteq i(l_1 \wedge \cdots \wedge l_n)$. Now, $i(\alpha \wedge l)$ is nonempty and $i(l_1 \wedge \cdots \wedge l_n)$ is a singleton. Then $i(\alpha \wedge l) = i(l_1 \wedge \cdots \wedge l_n)$. And $i(\alpha \wedge l) = i(\alpha) \Cup i(l) = \{a \cup b : a \in i(\alpha), b \in i(l)\}$. But as l is a literal, $i(l) = \{\{\iota(l)\}\}$ by semantic definition, and $i(l_1 \wedge \cdots \wedge l_n) = \{\{\iota(l_1), \ldots, \iota(l_n)\}\}$. Then $\{a \cup \iota(l) : a \in i(\alpha)\} = \{\iota(l_1), \ldots, \iota(l_n)\}$, and so $\iota(l) = \iota(l_k)$ for some $k \leq n$. Hence $w \models (l \hookrightarrow l_1) \vee \cdots \vee (l \hookrightarrow l_n)$. $\square$

A7 Soundness of IML

	Characteristic axiom scheme	**Corresponding semantic property**
(a)	$(\varphi \hookrightarrow \psi) \to ([\psi]\chi \to [\varphi]\chi)$	*Imaginative inclusion* If $i_w(\varphi) \subseteq i_w(\psi)$ and $wR_\varphi v$ then $wR_\psi v$
(b)	$([\varphi]\chi \wedge [\psi]\chi) \to [\varphi \vee \psi]\chi$	*Unconditionalisation* If $wR_{\varphi \vee \psi} v$ for some $v \in [\![\chi]\!]$, then $wR_\varphi u$ or $wR_\psi u$ for some $u \in [\![\chi]\!]$
(c)	$(\varphi \equiv \psi) \to (\varphi \leftrightarrow \psi)$	*Veracity* If $i_w(\varphi) = i_w(\psi)$ then $w \in [\![\varphi]\!]$ iff $w \in [\![\psi]\!]$
(d)	$[\varphi]\varphi$	*Success* If $wR_\varphi v$ then $v \models \varphi$
(e)	$[\varphi]\psi \to (\varphi \to \psi)$	*Weak centering* If $w \models \varphi$ then $wR_\varphi w$
(f)	$([\varphi]\psi \wedge [\varphi \wedge \psi]\chi) \to [\varphi]\chi$	*Cumulative transitivity* If $R_\varphi[w] \subseteq [\![\psi]\!]$ and $wR_\varphi v$ then $wR_{\varphi \wedge \psi} v$

Table 1: Axiom schema (a)–(f) with their corresponding semantic properties.

Proposition 65 (Soundness of each imaginative modal logic)**.** *Each axiom scheme (a)–(f) is valid on the class of frames with its corresponding semantic property.*

Proof. Pick any model $\mathcal{M} = (W, \{R_\varphi\}_{\varphi \in \mathcal{L}_{IML}}, V)$ of IML and any $w \in W$.

(a) Suppose $\mathcal{M}$ satisfies imaginative inclusion, $\mathcal{M}, w \models \varphi \hookrightarrow \psi$ and $\mathcal{M}, w \models [\psi]\chi$. Pick any $v \in W$ with $wR_\varphi v$. Then by imaginative inclusion, $wR_\psi v$. Then as $\mathcal{M}, w \models [\psi]\chi$, $\mathcal{M}, v \models \chi$.

(b) Suppose $\mathcal{M}$ satisfies unconditionalisation. It will be easier to use the axiom scheme $\langle \varphi \vee \psi \rangle \chi \to \langle \varphi \rangle \chi \vee \langle \psi \rangle \chi$, which is equivalent to $[\varphi]\chi \wedge [\psi]\chi \to [\varphi \vee \psi]\chi$, as the following chain of equivalences shows.

$$[\varphi]\neg\chi \wedge [\psi]\neg\chi \to [\varphi \vee \psi]\neg\chi \qquad \text{(Axiom scheme in } \Lambda)$$

$$\neg[\varphi \vee \psi]\neg\chi \to \neg[\varphi]\neg\chi \vee \neg[\psi]\neg\chi \qquad \text{(Propositional logic)}$$

$$\langle \varphi \vee \psi \rangle \chi \to \langle \varphi \rangle \chi \vee \langle \psi \rangle \chi \qquad \text{(Dual)}$$

Now suppose $\mathcal{M}, w \models \langle \varphi \vee \psi \rangle \chi$. Then for some $v \in W$ with $wR_{\varphi \vee \psi} v$ we have $\mathcal{M}, v \models \chi$. Then by unconditionalisation, $wR_\varphi u$ or $wR_\psi u$ for some $u \in W$ with $\mathcal{M}, u \models \chi$. That is, $\mathcal{M}, w \models \langle \varphi \rangle \chi$ or $\mathcal{M}, w \models \langle \psi \rangle \chi$, as required.

(c) Immediate.

(d) Suppose $\mathcal{M}$ satisfies success and pick any $v \in W$ with $wR_\varphi v$. Then by success, $\mathcal{M}, v \models \varphi$, as required.

(e) Suppose $\mathcal{M}$ satisfies weak centering and $\mathcal{M}, w \models \varphi \wedge [\varphi]\psi$. Then by weak centering, $wR_\varphi w$, so as $\mathcal{M}, w \models [\varphi]\psi$, we have $\mathcal{M}, w \models \psi$.

(f) Suppose $\mathcal{M}$ satisfies accumulation, $\mathcal{M}, w \models [\varphi]\psi$ and $\mathcal{M}, w \models [\varphi \wedge \psi]\chi$. Pick any $v \in W$ with $wR_\varphi v$. As $\mathcal{M}, w \models [\varphi]\psi$, by semantic definition, $R_\varphi[w] \subseteq [\![\psi]\!]$. Then as $wR_\varphi v$, by accumulation, $wR_{\varphi \wedge \psi} v$. So as $\mathcal{M}, w \models [\varphi \wedge \psi]\chi$, we have $\mathcal{M}, v \models \chi$.

□

A8 Completeness of IML

All lemmata (apart from the existence lemma and Proposition 41) required to prove completeness can be found in Blackburn et al. (2002, Chapter 4).

Lemma 66 (Existence lemma)**.** *Let Λ be a normal imaginative modal logic. For any Λ-MCS w, if $\langle \varphi \rangle \psi \in w$ then there is a Λ-MCS v such that $wR_\varphi v$ and $\psi \in v$.*

Proof. Suppose $\langle \varphi \rangle \psi \in w$ and consider $v_0 = \{\psi\} \cup \{\chi : [\varphi]\chi \in w\}$. Then v_0 is consistent. For suppose not. Then Λ proves $(\chi_1 \wedge \cdots \wedge \chi_n) \to \neg\psi$ for some $\chi_1, \ldots, \chi_n \in v_0$. Then Λ proves $([\varphi]\chi_1 \wedge \cdots \wedge [\varphi]\chi_n) \to [\varphi]\neg\psi$ using necessitation and K, so $[\varphi]\neg\psi \in w$, and by dual, $\neg\langle\varphi\rangle\psi \in w$. Then as w is a Λ-MCS, $\langle\varphi\rangle\psi \notin w$, contradicting the fact that $\langle\varphi\rangle\psi \in w$. As v_0 is consistent, by Lindenbaum's lemma $v_0 \subseteq v$ for some Λ-MCS v. Then $\psi \in v$. Also, $wR_\varphi v$ since, for any $[\varphi]\chi \in w$, $\chi \in v_0$ by definition, so $\chi \in v$ and we're done. $\square$

A9 Implicit hypotheticals: the coin flip

We begin by assigning imaginative contents to the following literals like so.

$$\begin{aligned} \iota(\text{bet T}) = \iota(\neg\text{bet H}) &= \mathcal{T} \\ \iota(\text{coin H}) = \iota(\neg\text{coin T}) &= \uparrow \\ \iota(\text{coin T}) = \iota(\neg\text{coin H}) &= \downarrow \end{aligned}$$

And we put $\iota(l) \neq \iota(l')$ for all distinct literals l, l' that are also distinct from the literals above. Recall that $c^+ = \{p : \iota(p) \in c\}$. Let $w = \{\text{bet T, coin T}\}$ be the actual world, $\textsc{implicit-hyp}_1 = \{\emptyset\}$ and $\textsc{implicit-hyp}_2 = \{\{\uparrow\}, \{\downarrow\}\}$. Also, $i(\text{bet T}) = \{\{\mathcal{T}\}\}$ since $\iota(\text{bet T}) = \mathcal{T}$. We calculate as follows.

$$\begin{gathered} \{\mathcal{T}\}^- = \{\text{bet H}\}, \quad \{\mathcal{T},\uparrow\}^- = \{\text{bet H, coin H}\}, \quad \{\mathcal{T},\downarrow\}^- = \{\text{bet H, coin H}\} \\ \{\mathcal{T}\}^+ = \{\text{bet T}\}, \quad \{\mathcal{T},\uparrow\}^+ = \{\text{bet T, coin H}\}, \quad \{\mathcal{T},\downarrow\}^+ = \{\text{bet T, coin T}\} \end{gathered}$$

$$\begin{aligned} wR^1_{\text{bet T}}v \text{ iff } & v = w^{c \cup h} \text{ for some } c \in ci(\text{bet T}) \text{ and } h \in \textsc{implicit-hyp}_1 \\ \text{iff } & v = w^{c \cup h} \text{ for some } c \in \{\{\mathcal{T}\}\} \text{ and } h \in \{\emptyset\} \\ \text{iff } & v = w^{\{\mathcal{T}\} \cup \emptyset} \\ \text{iff } & v = (w - \{\mathcal{T}\}^-) \cup \{\mathcal{T}\}^+ \\ \text{iff } & v = (\{\text{bet H, coin H}\} - \{\text{bet H}\}) \cup \{\text{bet T}\} \\ \text{iff } & v = \{\text{bet T, coin H}\} \end{aligned}$$

$$\begin{aligned} wR^2_{\text{bet T}}v \text{ iff } & v = w^{c \cup h} \text{ for some } c \in ci(\text{bet T}) \text{ and } h \in \textsc{implicit-hyp}_2 \\ \text{iff } & v = w^{c \cup h} \text{ for some } c \in \{\{\mathcal{T}\}\} \text{ and } h \in \{\{\uparrow\}, \{\downarrow\}\} \\ \text{iff } & v = w^{\{\mathcal{T}\} \cup \{\uparrow\}} \text{ or } v = w^{\{\mathcal{T}\} \cup \{\downarrow\}} \\ \text{iff } & v = (w - \{\mathcal{T},\uparrow\}^-) \cup \{\mathcal{T},\uparrow\}^+ \text{ or } v = (w - \{\mathcal{T},\downarrow\}^-) \cup \{\mathcal{T},\downarrow\}^+ \\ \text{iff } & v = (\{\text{bet H, coin H}\} - \{\text{bet H, coin H}\}) \cup \{\text{bet T, coin H}\} \\ \text{or } & v = (\{\text{bet H, coin H}\} - \{\text{bet H, coin H}\}) \cup \{\text{bet T, coin T}\} \\ \text{iff } & v = \{\text{bet T, coin H}\} \text{ or } v = \{\text{bet T, coin T}\}. \end{aligned}$$

Bibliography

Maria Aloni. Free choice, modals, and imperatives. *Natural Language Semantics*, 15(1): 65–94, Mar 2007. doi:10.1007/s11050-007-9010-2.

Maria Aloni and Ivano Ciardelli. A semantics for imperatives. Unpublished manuscript, 2011. URL `https://pdfs.semanticscholar.org/9cdd/fb56130329b403ff5240dbeb5ef3e3f2ccaa.pdf`.

Luis Alonso-Ovalle. *Disjunction in alternative semantics*. PhD thesis, University of Massachusetts Amherst, 2006. URL `http://people.linguistics.mcgill.ca/~luis.alonso-ovalle/papers/alonso-ovalle-diss.pdf`.

Luis Alonso-Ovalle. Counterfactuals, correlatives, and disjunction. *Linguistics and Philosophy*, 32(2):207–244, Apr 2009. doi:10.1007/s10988-009-9059-0.

Alan Ross Anderson. A note on subjunctive and counterfactual conditionals. *Analysis*, 12(2):35–38, 1951. doi:10.1093/analys/12.2.35.

Jon Barwise and John Perry. *Situations and attitudes*. MIT Press, 1983.

Francesco Berto. Impossible worlds and the logic of imagination. *Erkenntnis*, 82(6):1277–1297, Dec 2017. doi:10.1007/s10670-017-9875-5.

Francesco Berto. Aboutness in imagination. *Philosophical Studies*, 175(8):1871–1886, Aug 2018. ISSN 1573-0883. doi:10.1007/s11098-017-0937-y. URL `https://doi.org/10.1007/s11098-017-0937-y`.

Francesco Berto, Rohan French, Graham Priest, and David Ripley. Williamson on counterpossibles. *Journal of Philosophical Logic*, 47(4):693–713, Aug 2018. doi:10.1007/s10992-017-9446-x.

Patrick. Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 2002. doi:10.1017/CBO9781107050884.

Matthew Braham and Martin Van Hees. An anatomy of moral responsibility. *Mind*, 121 (483):601–634, 7 2012. doi:10.1093/mind/fzs081.

Robert Brandom. *Articulating Reasons: an Introduction to Inferentialism*. Harvard University Press, 2000. URL `www.hup.harvard.edu/catalog.php?isbn=9780674006928`.

Rachael Briggs. Interventionist counterfactuals. *Philosophical Studies*, 160(1):139–166, Aug 2012. doi:10.1007/s11098-012-9908-5.

Berit Brogaard and Joe Salerno. Why counterpossibles are non-trivial. In Vincent Hendricks, editor, *Synthese*. Springer, forthcoming.

Ruth MJ Byrne. *The Rational Imagination: How people create alternatives to reality*. MIT press, 2007. URL `https://mitpress.mit.edu/books/rational-imagination`.

C. S. Lewis. *Rehabilitations and other essays*. Oxford university press, 1939.

Nate Charlow. The meaning of imperatives. *Philosophy Compass*, 9(8):540–555, 2014. doi:10.1111/phc3.12151.

Brian F. Chellas. Basic conditional logic. *Journal of Philosophical Logic*, 4(2):133–153, May 1975. doi:10.1007/BF00693270.

Ivano Ciardelli, Jeroen Groenendijk, and Floris Roelofsen. Inquisitive semantics: A new notion of meaning. *Language and Linguistics Compass*, 7(9):459–476, 2013. doi:10.1111/lnc3.12037.

Ivano Ciardelli, Jeroen Groenendijk, and Floris Roelofsen. *Inquisitive Semantics*. Oxford Surveys in Semantics and Pragmatics. Oxford University Press, 2018a.

Ivano Ciardelli, Linmin Zhang, and Lucas Champollion. Two switches in the theory of counterfactuals. *Linguistics and Philosophy*, Jun 2018b. doi:10.1007/s10988-018-9232-4.

Hannah Clark-Younger. *Imperatives and Logical Consequence*. PhD thesis, University of Otago, 2014. URL `https://ourarchive.otago.ac.nz/bitstream/handle/10523/5039/ClarkYoungerHannah2014PhD.pdf?sequence=1`.

Edmund M. Clarke and E. Allen Emerson. Design and synthesis of synchronization skeletons using branching time temporal logic. In Dexter Kozen, editor, *Logics of Programs*, pages 52–71. Springer, 1982. doi:10.1007/BFb0025774.

Lewis G. Creary and Christopher S. Hill. Review of Counterfactuals by David Lewis. *Philosophy of Science*, 42(3):341–344, 1975. doi:10.1086/288656.

M. J. Cresswell. A Henkin completeness theorem for T. *Notre Dame Journal of Formal Logic*, 8(n/a):186, 1967. doi:10.1305/ndjfl/1093956082.

Brian Ellis, Frank Jackson, and Robert Pargetter. An objection to possible-world semantics for counterfactual logics. *Journal of Philosophical Logic*, 6(1):355–357, Jan 1977. doi:10.1007/BF00262069.

Kit Fine. Critical notice (review of Lewis 1973b). *Mind*, 84(1):451–458, 1975. doi:10.1093/mind/LXXXIV.1.451.

Kit Fine. Counterfactuals without possible worlds. *Journal of Philosophy*, 109(3):221–246, 2012a. doi:10.5840/jphil201210938.

Kit Fine. A difficulty for the possible worlds analysis of counterfactuals. *Synthese*, 189 (1):29–57, 2012b. doi:10.5840/jphil201210938.

Kit Fine. Truth-maker semantics for intuitionistic logic. *Journal of Philosophical Logic*, 43 (2):549–577, Jun 2014. doi:10.1007/s10992-013-9281-7.

Kit Fine. Angellic content. *Journal of Philosophical Logic*, 45(2):199–226, 2016. doi:10.1007/s10992-015-9371-9.

Kit Fine. Truthmaker semantics. In *A Companion to the Philosophy of Language*, chapter 22, pages 556–577. Wiley-Blackwell, 2017. doi:10.1002/9781118972090.ch22.

Kai von Fintel and Sabine Iatridou. A modest proposal for the meaning of imperatives. *Modality across syntactic categories*, pages 288–319, 2017. URL `http://web.mit.edu/fintel/fintel-iatridou-2015-modest.pdf`.

Gottlob Frege. On sense and reference. *The Philosophical Review*, 57(3):209–230, 1948. doi:10.2307/2181485.

David Galles and Judea Pearl. An axiomatic characterization of causal counterfactuals. *Foundations of Science*, 3(1):151–182, Jan 1998. doi:10.1023/A:1009602825894.

Nelson Goodman. The problem of counterfactual conditionals. *Journal of Philosophy*, 44

(5):113–128, 1947. doi:10.2307/2019988.

H Paul Grice. *Studies in the Way of Words*. Harvard University Press, 1991.

Joseph Halpern. Axiomatising causal reasoning. *Journal of Arti cial Intelligence Research*, 12, 2000. doi:10.1613/jair.648.

Eric Hiddleston. A causal theory of counterfactuals. *Noûs*, 39(4):632–657, 2005. doi:10.1111/j.0029-4624.2005.00542.x.

Laurence R. Horn. Metalinguistic negation and pragmatic ambiguity. *Language*, 61(1): 121–174, 1985. doi:10.2307/413423.

Sabine Iatridou. The grammatical ingredients of counterfactuality. *Linguistic Inquiry*, 31 (2):231–270, 2000. doi:10.1162/002438900554352.

Sabine Iatridou and Kai von Fintel. Transparent wishes. Lecture slides, 2017. URL `http://web.mit.edu/fintel/fintel-iatridou-2017-x-desires-slides-ipac.pdf`.

Michela Ippolito. *Subjunctive Conditionals: A Linguistic Analysis*. MIT Press, 2013.

Daniel Lassiter. Complex sentential operators refute unrestricted simplification of disjunctive antecedents. *Semanitcs and Pragmatics*, 11, 2018. doi:10.3765/sp.11.9.

Edward John Lemmon and Dana Scott. *An Introduction to Modal Logic*, volume 11 of *American Philosophical Quarterly Monograph Series*. Basil Blackwell, 1977. Written by E. J. Lemmon in collaboration with D. Scott.

David Lewis. Completeness and decidability of three logics of counterfactual conditionals1. *Theoria*, 37(1):74–85, 1971. doi:10.1111/j.1755-2567.1971.tb00061.x.

David Lewis. General semantics. In Donald Davidson and Gilbert Harman, editors, *Semantics of Natural Language*, pages 169–218. Springer Netherlands, Dordrecht, 1972. doi:10.1007/978-94-010-2557-7_7.

David Lewis. Causation. *Journal of Philosophy*, 70(17):556–567, 1973a. doi:10.2307/2025310.

David Lewis. *Counterfactuals*. John Wiley & Sons, 1973b.

David Lewis. Scorekeeping in a language game. *Journal of Philosophical Logic*, 8(1):339–359, 1979. doi:10.1093/0195032047.001.0001.

David Lewis. *Counterfactuals and Comparative Possibility*, pages 57–85. Springer, 1981. doi:10.1007/978-94-009-9117-0_3.

David Makinson. On some completeness theorems in modal logic. *Mathematical Logic Quarterly*, 12(1):379–384, 1966. doi:10.1002/malq.19660120131.

Thomas McKay and Peter van Inwagen. Counterfactuals with disjunctive antecedents. *Philosophical Studies*, 31(5):353–356, May 1977. doi:10.1007/BF01873862.

Daniel Nolan. Impossible worlds: A modest approach. *Notre Dame Journal of Formal Logic*, 38(4):535–572, 1997. doi:10.1305/ndjfl/1039540769.

Donald Nute. Counterfactuals and the similarity of words. *The Journal of Philosophy*, 72 (21):773–778, 1975. doi:10.2307/2025340.

Donald Nute. Conversational scorekeeping and conditionals. *Journal of Philosophical Logic*, 9(2):153–166, May 1980. doi:10.1007/BF00247746.

Peter Pagin and Dag Westerståhl. Compositionality I: Definitions and variants. *Philosophy Compass*, 5(3):250–264, 2010. doi:10.1111/j.1747-9991.2009.00228.x.

Josh Parsons. Command and consequence. *Philosophical Studies*, 164(1):61–92, May 2013. doi:10.1007/s11098-013-0094-x.

Huw Price and Brad Weslake. The time-asymmetry of causation. In Helen Beebee, Peter

Menzies, and Christopher Hitchcock, editors, *The Oxford Handbook of Causation*. Oxford University Press, 2009. URL http://philsci-archive.pitt.edu/id/eprint/4475.

Graham Priest. *An introduction to non-classical logic: From if to is*. Cambridge University Press, 2008.

Willard Van Orman Quine. *Word and object*. MIT press, 2013.

Paolo Santorio. Interventions in premise semantics. *Philosophers' Imprint*, 2016.

Katrin Schulz. Non-deictic tenses in conditionals. In *Semantics and Linguistic Theory*, volume 18, pages 694–710, 2008. doi:10.3765/salt.v18i0.2498.

Katrin Schulz. "If you'd wiggled A, then B would've changed". *Synthese*, 179(2):239–251, Mar 2011. doi:10.1007/s11229-010-9780-9.

Katrin Schulz. Fake perfect in X-marked conditionals. In *Semantics and Linguistic Theory*, volume 27, pages 547–570, 2017. doi:10.3765/salt.v27i0.4149.

Robert Stalnaker. A theory of conditionals. In *Ifs*, pages 41–55. Springer, 1968.

William B. Starr. A uniform theory of conditionals. *Journal of Philosophical Logic*, 43(6): 1019–1064, Dec 2014. doi:10.1007/s10992-013-9300-8.

Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974. doi:10.1126/science.185.4157.1124.

Dirk Van Dalen. *Logic and Structure*. Springer, 1994. doi:10.1007/978-3-662-02962-6.

Ken Warmbrōd. Counterfactuals and substitution of equivalent antecedents. *Journal of Philosophical Logic*, 10(2):267–289, May 1981. doi:10.1007/BF00248853.

Malte Willer. Simplifying counterfactuals. In Thomas Brochhagen, Floris Roelofsen, and Nadine Theiler, editors, *Proceedings of the 20th Amsterdam Colloquium*, pages 428–437, 2015. URL http://philosophy.uchicago.edu/faculty/files/willer/Simplifying%20Counterfactuals.pdf.

Malte Willer. Simplifying with free choice. *Topoi*, 37(3):379–392, Sep 2018. doi:10.1007/s11245-016-9437-5.

Timothy Williamson. Counterpossibles in metaphysics. *Philosophical Fictionalism*, 2017.

Timothy Williamson. Counterpossibles. *Topoi*, 37(3):357–368, Sep 2018. doi:10.1007/s11245-016-9424-x.