WHAT DOES A CONDITIONAL MEAN?

It is customary to characterise conditionals as compound linguistic expressions consisting of two sentences conjoined by a connective “if.” Roughly speaking, the *if*-clause, also referred to as the *antecedent* or *protasis*, expresses a condition under which the main clause of the conditional sentence, that is its *consequent* or *apodosis*, is meant to hold. A paradigmatic conditional is hence a sentence of the form: “If φ, (then) ψ,” or, alternatively, “ψ if φ,” like, for instance, the following sentences:

(12) a. A book is not eligible for the Man Booker Prize if it has not been originally written in English.
    
    b. If Francisco Goya did not paint the black paintings himself, his son Javier must have painted them.
    
    c. If Alice Munro had not been awarded the Nobel Prize in Literature in 2013, someone else would have received it.
    
    d. If Maria Skłodowska-Curie had not married a Frenchman, people would not tend to think that she was French.

Of the above sentences, (12a) and (12b) are traditionally referred to as *indicative conditionals* (or *indicatives*, for short), whereas (12c) and (12d) are called *subjunctive conditionals* (or *subjunctives*). To illustrate the semantic difference between indicatives and subjunctives, various authors typically invoke the following two sentences due to Adams (1970):

(13) a. If Oswald did not kill Kennedy, someone else did.
    
    b. If Oswald had not killed Kennedy, someone else would have.

Here, (13a) is an indicative and (13b) is a subjunctive. Subjunctive conditionals are frequently counterfactual and vice versa, yet the two terms are not interchangeable. The term “subjunctive conditional” should be understood as indicating a grammatical category, while “counterfactual” is a semantic notion. A conditional is counterfactual when it presupposes the falsehood of its antecedent, and not all subjunctives do that. To give an example, the sentence:
(14) If he were to marry her, he would have to move to Finnland.
is a subjunctive conditional, yet it can be asserted by a speaker
for whom the antecedent is an open possibility. At the same time,
one might assert an indicative:

(15) If Denmark is ruled by a king, it is a kingdom.
even if they know that the Kingdom of Denmark is not ruled by
a king, but by a queen, if in the given context it does not matter
who the actual ruler is. Such a conditional could be asserted, for
instance, as an instance of an inference from “a country is ruled
by a king” to “a country is a kingdom.” Given that this work
is mostly concerned with indicative conditionals, the unqualified
term “conditionals” or “conditional sentences” will henceforth
refer to indicatives.

The last few decades witnessed a growing interest among re-
searchers of various backgrounds in the issues related to condi-
tionals. Consequently, countless theories trying to account for the
meaning of conditional sentences have been developed. It would
be pointless, if not utterly impossible, to even try to discuss them
all in any detail, especially given that many outstanding works
reviewing the available literature have been published in recent
years. To name just a few, Bennett (2003) and Edgington (2014)
offer comprehensive guides through the philosophical issues re-
lated to conditionals. Sanford (1989), by contrast, takes a historical
perspective in his presentation. Discussions of various approaches
towards conditional logic can be found in Nute and Cross (2002)
or Arló-Costa (2007), while Evans and Over (2004) provide a thor-
ough analysis of psychological results concerning the interpreta-
tion of conditional sentences. Of more recent works, Douven (in
press) explores the epistemological issues raised by conditional
sentences, demonstrating additionally the benefits of applying
both formal and empirical methods to philosophical analysis.

Instead, to prepare the grounds for the presentation of my own
results, I will focus on some of the most distinctive and prob-
lematic features of two classes of approaches towards conditional
sentences that do not contest their propositionality. First, I will
review strengths and flaws of a truth-functional account of con-
ditionals, that is, the material account, according to which “If \( \varphi \),
(then) \( \psi \)” is equivalent to an inclusive disjunction of \( \neg \varphi \) and \( \psi \).
Second, I will discuss truth-conditional theories of conditionals
inspired by the Ramsey Test, focusing on the possible world se-
manics developed by Stalnaker (1968). But before I move on to
analysing the above mentioned accounts, let me touch upon the issue of what a conditional sentence actually is.

2.1 INTERLUDE: CONDITIONALS AND “IFS”

Even though a prototypical conditional sentence can be characterised by the presence of a connective “if,” it would be wrong, however tempting, to conclude that studying conditionals is nothing more than studying the function or the meaning of the word “if” alone. Associating one with the other seems natural especially from the perspective of native English speakers, but one should not forget that English is not necessarily the most representative language in the world. Any claims about language that are intended as universal, or at least as more general than statements about particular features of a specific language, cannot be based solely on linguistic data drawn from a single source.

Even if we look into very limited cross-linguistic data from, for instance, languages relatively closely related to English like other European languages, we can easily find evidence in favour of a separate treatment of conditional sentences and sentences with if-clauses. First and foremost, there are languages in which English if can be translated in more than one way, depending on the linguistic or extralinguistic context. In Polish, for instance, a subordinate clause of an indicative conditional can be introduced by means of jeśli or jeżeli. The sentence:

(16) Jeśli Beata wie, to musi się martwić.

If Beata knows then must worry.

If Beata knows, she must be worried.

is roughly equivalent to:

(17) Jeżeli Beata wie, to musi się martwić.

There is no evident semantic difference between the two Polish indicative ifs. The word jeżeli is perhaps more formal, but one could argue that the choice between jeśli and jeżeli amounts to something more than a matter of style. Jeżeli as longer and thus

1 In fact, some seminal works devoted broadly to conditionals and conditional reasoning are simply titled “If” or “Ifs” (Evans and Over 2004; Harper et al. 1981, respectively).

2 To be precise, there is no semantic difference that I, as a native Polish speaker, am able to observe. I am also not aware of any corpus-driven or experimental research on differences between Polish jeśli and jeżeli.
less economical seems to be most felicitous when a speaker wants
to stress that what is being said is hypothetical, or to draw an
interlocutor’s attention to the content of the antecedent. For this
reason, (17) may in some contexts sound somewhat emotionally
loaded while (16) would remain entirely neutral.

By contrast, a subjunctive conditional in Polish involves yet an-
other connective that is translated as if into English, namely gdyby:

(18) Gdyby Beata wiedziała, to by się martwiła.

    If Beata had known then would have worried.

    If Beata had known, she would have worried.

Furthermore, if is not the only English connective linking the
main and the subordinate clauses of a conditional. On the basis of
an extensive study of linguistic corpora, Declerck and Reed (2001)
note that conditional clauses can be also introduced by means of
expressions like unless, provided that, in case, supposing, assuming
and many others, including connectives typically associated with
temporal clauses like when or as soon as. Though it usually implies
factuality, when can have a clearly conditional connotation, e.g.:

(19) I will stop nagging you when you start doing what you’ve
    promised. (Declerck and Reed 2001, p. 32)

Moreover, Declerck and Reed (2001, p. 33) claim that in cases like
the following:

(20) a. Children are orphans when their parents are dead.

    b. Children are orphans if their parents are dead.

when- and if-clauses can be used interchangeably.3

Polish is additionally equipped with connectives such as skoro
and jak that seem to have both temporal and conditional connota-
tions. Asserting a conditional with a jak-clause seems to indicate
that the speaker’s degree of belief that the antecedent holds is
rather high, although not as high as when kiedy or gdy (which
can be directly translated into English as when) are used. By con-
trast, skoro, when it is used in a conditional (that is, not purely
temporal) clause, seems roughly equivalent to English given that
or provided that.

---

3 See also Elder (2012) for a corpus-driven exploration of different ways a condi-
tional can be expressed in English.
Another piece of evidence in favour of a separate analysis of a conditional on the one hand, and of the connective, on the other hand, is the fact that it is not necessary for a conditional to involve any connective at all:

(21) No broccoli, no dessert.

The above example clearly expresses a conditional dependency. However, one could argue that (21) is not really a sentence, but, for instance, an abbreviation that can be developed into a full sentence along the following lines:

(22) If you do not eat your broccoli, you will not get the dessert.

Nevertheless, the constructions with so-called zero-conjunction and inversion can constitute full-fledged conditional sentences (for a more detailed analysis of these, see Declerck and Reed 2001):

(23) a. Had she told him earlier, he would not have been so furious.

b. Should someone ring, tell them I’ll be at the office till six.
   (Declerck and Reed 2001, p. 27)

c. Were he to try that again, I’d go to the police. (ibid.)

A similar phenomenon can be also observed in Polish:

(24) a. Odwiedzisz mnie, to sam zobaczysz.
   You will visit me then yourself you will see.
   *If you visit me, then you will see for yourself.*

b. Porozmawiałbyś z nim, to by zrozumiał.
   You would talk to him then he would understand.
   *Had you talked to him, he would have understood.*

Yet another reason to disentangle the analysis of conditionals from the analysis of if is the presence of this connective in sentences whose conditionality is questionable. One could argue, for instance, that the following sentences:

(25) a. If this is true, I’m a Dutchman.

b. If that’s Jack who wrote this essay, I am a monkey’s uncle.

are just a fanciful way to say, respectively:
(26) a. This cannot possibly be true.
   
   b. Jack could not have possibly written this essay.

In principle, however, (25a) and (25b) can be seen as proper conditionals that simply convey somewhat unusual thoughts, namely, that supposing their antecedents leads to ridiculous conclusions. Sentences belonging to the class of so called speech-act conditionals constitute perhaps a more compelling example of linguistic constructions with “if” whose conditionality can be contested, for instance:

(27) a. If you are hungry, there are biscuits on the table.
   
   b. If you really must know, Bill did not come.

In (27a), clearly, the content of the consequent is asserted unconditionally: the biscuits are on the table regardless whether the interlocutor is hungry or not. The only purpose the if-clause of this sentence seems to serve is of a pragmatic kind. It directs a hearer’s attention to the asserted information or indicates when that information is relevant for the hearer. In (27b), similarly, the antecedent is not a condition under which the consequent is supposed to hold, but rather a remark suggesting that what follows is said somewhat reluctantly. In more general terms, what is conditionally modified by the content of an if-clause in the case of speech-act conditionals is the act of asserting the main clause, not its content (Dancygier and Sweetser 2005, p. 113).

Although the interpretation of the above reported phenomena is likely to remain a matter of some controversy—a controversy which is not my ambition here to resolve—I believe that they constitute a good enough reason not to think of if as being all there is to the analysis of conditional sentences. That being said, the example sentences I will use to illustrate the theory proposed in this dissertation will mostly be sentences with if-clauses, as those are the most typical cases of conditionals. It is nonetheless important to bear in mind that what signifies a conditional sentence is not its particular surface structure, or more specifically, a particular connective.

2.2 The Ideal: A Truth-Functional Account

Conditionals are complex linguistic expressions. They are sentences compounded of two simpler sentences, which can be com-
plex themselves, usually (but not necessarily, cf. section 2.1) con-
joined by means of a connective “if.” A noble tradition cultivated
in semantics and philosophy of language teaches us to analyse
meanings of complex expressions as functions of the meanings
of their constituents and the way they are syntactically combined
(see, e.g., Partee 1984; Janssen 1997). This idea, known as the Prin-
ciple of Compositionality, derives from writings of Gottlob Frege
who realised that the immense productivity of language can only
be accounted for by the existence of some mechanism allowing us
to decode the correspondence between the syntactic structure and
the structure of the thought it expresses. As he writes in “Com-
pound Thoughts”:

It is astonishing what language can do. With a few
syllables it can express an incalculable number of
thoughts, so that even a thought grasped by a ter-
restrial being for the very first time can be put into a
form of words which will be understood by someone
to whom the thought is entirely new. This would be
impossible, were we not able to distinguish parts in
the thoughts corresponding to the parts of a sentence,
so that the structure of the sentence serves as the im-
age of the structure of the thoughts. (Frege 1963)

In Fregean philosophy, both the meaning (Sinn) and the reference
(Bedeutung) of a complex expression are compositional. As the
reference of a sentence is its truth value, where \( \varphi \) and \( \psi \) are sen-
tences and \( \ast \) is some binary sentential operator conjoining them,
the truth value of \( \varphi \ast \psi \) depends on the truth values of \( \varphi \) and \( \psi \)
as well as on the structure of the whole expression determined by
the operator \( \ast \). Ideally, this dependency is functional, that is, the
truth value of a complex sentence is a function of the truth values
of its parts. Hence, a theory of conditionals in which intuitions
articulated in the above quote are realised in the simplest and
perhaps the most elegant way is the so called material account,
sometimes referred to as a horseshoe analysis of a conditional
due to the convention of using the sign “\( \supset \)” as a material condi-
tional connective.\(^5\)

The material conditional, \( \varphi \supset \psi \), inherited its name after
the notion of material implication introduced by Bertrand Russell

---

4 Though traditionally attributed to Frege and indubitably in the spirit of his late
works, there is no clear evidence that the Principle of Compositionality has been
endorsed by Frege as a principle. See Janssen (1997) for a discussion of this issue.
5 This is the convention I am going to follow from now on.
and Alfred North Whitehead in *Principia Mathematica* (1962, p. 7; see also Sanford 1989, pp. 50-52). However, the first philosopher to whom the truth-functional analysis of a conditional can be attributed is a stoic philosopher, Philo of Megara (Sanford 1989, pp. 15-23), whence the term “Philonian conditional” is also to be encountered in the literature. On this account, the semantics of a natural language conditional is identical to that of an implication as defined in classical logic. Of the four possible ways we can assign the truth values, \{0, 1\}, to the two constituents, \(\varphi\) and \(\psi\), only one results in the implication being false, namely, when the antecedent is true but the consequent is false. In other words, a material conditional is true if and only if either its antecedent is false, or its consequent is true:

\[
\varphi \supset \psi \equiv \neg \varphi \lor \psi
\]  

(1)

Analogously to other classical logic formulas, the meaning of a conditional is exhausted by the following truth table:

<table>
<thead>
<tr>
<th>\varphi</th>
<th>\psi</th>
<th>\varphi \supset \psi</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

One can clearly see that this is a truth-functional interpretation: the truth value of a conditional is determined by the truth values of its antecedent and consequent alone, exactly as it is in the case of logical conjunctions and disjunctions.

Truth-conditionality is not only a theoretical virtue by itself. What is more, one of the strongest arguments in favour of the material account is an immediate consequence of its truth-functionality, namely, it allows us to infer a conditional, “If \(\neg \varphi, \psi\),” from the disjunction, \(\varphi \lor \psi\). The *or-to-if* inference is not merely logically valid, but it also seems intuitively appealing and relatively prevalent in our ordinary everyday reasoning. For instance, if I do not remember whether I left my copy of Lewis’s *Counterfactuals* at home or in the office, but I am quite sure that the book must be in one of these places, I instantly believe that if the book is not at home, it is in the office:

28  a. Either the book is at home or it is in the office.

b. *Therefore*, if the book is not at home, it is in the office.
The above inference appears so natural that validating it would seem a highly desirable feature of a theory of natural language conditionals (however, we will discuss this allegedly uncontroversial issue in section 3.4). The material account renders the above inference valid (Stalnaker 1975). More importantly still, as demonstrated by Edgington (1995, 2014), it is also the only account that does that. For let us assume that \( \varphi \lor \psi \) is known, or in other words that we know that \( \varphi \) and \( \psi \) cannot be both false. To see that this is sufficient for us to infer a material conditional \( \neg \varphi \supset \psi \), but not a non-truth functional conditional, denoted here by \( \neg \varphi \rightarrow \psi \), let us consider the following table:

<table>
<thead>
<tr>
<th>( \varphi )</th>
<th>( \psi )</th>
<th>( \varphi \lor \psi )</th>
<th>( \neg \varphi \supset \psi )</th>
<th>( \neg \varphi \rightarrow \psi )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0 or 1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0 or 1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0 or 1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

It is worth noting that a non-truth-functional interpretation of a conditional is usually represented as departing from the material interpretation only in the cases where the antecedent is false, which in the above case would be the first and the second row of the table. This is because Stalnaker’s truth-conditional semantics is the best known alternative to the truth-functional account, and on this interpretation, as we will see later in this chapter, a conditional is true whenever its antecedent and the consequent are true. This is not only an unnecessary feature of a truth-conditional semantics, but also, as I will argue, one of the weaknesses of Stalnaker’s account. Nevertheless, for the present argument to go through, that is, for \( \rightarrow \) to be a non-truth-functional operator, it is sufficient that there is just one way to assign truth-values to the constituents, \( \varphi \) and \( \psi \), such that \( \rightarrow \) does not return a unique value.

Knowing that at least one of \( \varphi \) and \( \psi \) is true allows us to eliminate the bottom row which represents \( \neg \varphi \land \neg \psi \), which is incompatible with our knowledge. One can clearly see that \( \neg \varphi \supset \psi \) is true whenever the disjunction is true, thus \( \varphi \lor \psi \) entails the material conditional. By contrast, eliminating the bottom row of the table does not lead us to any certainty that \( \neg \varphi \rightarrow \psi \) is true. For all we know, the non-truth-functional conditional can still be false. \( \varphi \rightarrow \psi \) is therefore not entailed by the disjunction of \( \varphi \) and \( \psi \).
As conditionals play a vital role in reasoning, be it in science, mathematical proofs, or in our everyday decision making and planning, being able to demarcate correct and incorrect inferences or good and bad arguments is of utmost importance for our existence. One of the advantages of the material account is that it allows us to apply classical logic to evaluate arguments articulated in natural language. More precisely, it allows us to recognise logically valid and logically invalid arguments just on the basis of their form. If there is such a truth value assignment that results in true premises but a false conclusion, the argument form is logically invalid. Otherwise, the argument is logically valid.

Apart from the or-to-if inference discussed above, the most important argument forms involving conditional sentences are four elimination inferences: Modus Ponens (MP), Modus Tollens (MT), Affirmation of the Consequent (AC) and Denial of the Antecedent (DA). In each of them, a conditional, $\varphi \supset \psi$, acts as a major premise, and one of its constituents, a categorical $\varphi$ or $\psi$, as a minor premise. The conclusion is again a categorical, $\varphi$ or $\psi$, hence a conditional is in this type of argument being eliminated.

Of the four aforementioned argument forms, the first two are logically valid, and the last two logically invalid. Moreover, the valid forms seem intuitively appealing, in a sense that, at least prima facie, they seem to hold for the ordinary language conditional, too. Modus Ponens:

$$
\varphi \supset \psi, \varphi
\over \psi
$$

is an inference pattern that is often invoked in our everyday thought processes or discussions, for instance:

(29) a. If Paulina has been to Ljubljana, then she has been to Slovenia.
    Paulina has been to Ljubljana.
    Therefore, Paulina has been to Slovenia.

b. If the Netherlands is ruled by a king, then it is a monarchy.
    The Netherlands is ruled by a king.
    Therefore, The Netherlands is a monarchy.

c. If Alex is a vegetarian, then he doesn’t eat meat.
    Alex is a vegetarian.
    Therefore, Alex doesn’t eat meat.
Moreover, it seems to play a critical role in our everyday deliberations, which makes it an important component of planning and decision making:

(30) a. If you want to become a professional cellist, you must practice regularly.
   You want to become a professional cellist.
   Therefore, you must practice regularly.

b. If I don’t want to overpay, I should book my flight in advance.
   I don’t want to overpay.
   Therefore, I should book my flight in advance.

c. If you are interested in conditionals, you should read Jonathan Bennett’s book.
   You are interested in conditionals.
   Therefore, you should read Jonathan Bennett’s book.

Data from countless reasoning experiments also show that MP is relatively easy and usually endorsed by the participants. In fact, it is more frequently endorsed than any other inference form, including Modus Tollens (see Evans and Over 2004, pp. 46-52, and references there). It might be partly due to the fact that it is compatible with both a conjunctive and a biconditional interpretation of the conditional attributed to some participants, in particular, to children, adolescents, and cognitively less able adults (Barrouillet et al. 2000).

By contrast, Affirmation of the Consequent (AC):

\[
\varphi \supset \psi, \psi \\
\therefore \varphi
\]

is not a valid argument, yet its endorsement rates in different experiments range from 23 to 75% (Evans and Over 2004, p. 51).

(31) a. If Paulina has been to Ljubljana, then she has been to Slovenia.
   Paulina has been to Slovenia.
   Therefore, Paulina has been to Ljubljana.

b. If Ukraine is ruled by a king, then it is a monarchy.
   Ukraine is a monarchy.
   Therefore, Ukraine is ruled by a king.

c. If Alex is a vegetarian, he doesn’t eat meat.
   Alex doesn’t eat meat.
   Therefore, Alex is a vegetarian.
One can easily see that the above inferences are flawed. Paulina might have been to, for instance, the Slovenian town of Bled and never visited the country’s capital, and a monarchy can also be ruled by a queen. Interestingly, the conclusion of (31c), though the argument is still invalid, seems fairly appealing. The antecedent of the conditional given as the major premise may be perceived as sufficient for the truth of the consequent, which facilitates biconditional interpretation (Thompson 1994; Evans and Over 2004, p. 96). It might be the case that people tend to assert a conditional when a biconditional is equally acceptable: “Alex is a vegetarian if and only if he doesn’t eat meat.” English does not seem to be equipped with a single word connective that could be used to express a biconditional. The phrase “if and only if” seems to belong to a mathematical jargon rather than to an ordinary language. The phrases such as “precisely if” or “just in case” do not seem to be used frequently either. English speakers may prefer to assert just one of the two conditionals entailed by a biconditional they actually believe, especially if only one of them is relevant in the context of a conversation. In consequence, however, this might lead to what seems to be an erroneous practice of reading biconditional statements into conditional assertions, and, accordingly, to false conclusions.

Similarly, the third elimination inference, Denial of the Antecedent (DA):

$$\varphi \supset \psi, \neg \varphi$$

$$\therefore \neg \psi$$

is invalid, but sometimes convincing, and hence endorsed (Evans and Over 2004, p. 46, report 19-73% endorsement rates for DA inferences across various studies).

(32) a. If Paulina has been to Ljubljana, then she has been to Slovenia.
Paulina hasn’t been to Ljubljana.
Therefore, Paulina has’t been to Slovenia.

b. If the Netherlands is ruled by a king, then it is a monarchy.
The Netherlands is not ruled by a king.
Therefore, The Netherlands is not a monarchy.

---

6 This is also true for, e.g., Polish, Dutch, or German, and, presumably, many other languages.
c. If Alex is a vegetarian, then he doesn’t eat meat.
Alex is not a vegetarian.
Therefore, Alex eats meat.

One can easily imagine someone accepting, for instance, (32b) as a correct inference just because they did not realise the possibility of a queen-ruled monarchy. (32c) can again be interpreted so that the conclusion, “Alex eats meat” is true. The endorsement of DA inferences is also linked to a biconditional interpretation of “if..., then...” statements typical for adolescents, though also shown by some adults (Barrouillet et al. 2000). Evans and Over (2004) point out, however, that the biconditional pattern of responses does not necessarily indicate a truth-functional interpretation of conditional sentences:

It can simply indicate a superficial reading that p and q go together. If you have one, you have the other; if you do not have one, you don not have the other (p. 52).

This might also be the reason why Modus Tollens:

\[
\frac{\varphi \supset \psi \quad \neg \psi}{\neg \varphi}
\]

is not as frequently endorsed by the participants of the reasoning experiments as MP. Even though it is a valid inference rule, the endorsement rates across various studies have been reported to range from 14 to 81% (Evans and Over 2004, p. 46). At least at first glance, MT seems as intuitively appealing as MP:

(33) a. If Paulina has been to Ljubljana, she has been to Slovenia.
Paulina hasn’t been to Slovenia.
Therefore, Paulina hasn’t been to Ljubljana.

b. If Ukraine is ruled by a king, it is a monarchy.
Ukraine is not a monarchy.
Therefore, Ukraine is not ruled by a king.

c. If Alex is a vegetarian, he doesn’t eat meat.
Alex eats meat.
Therefore, Alex is not a vegetarian.

Yet it seems to be more difficult and more cognitively demanding than MP (see, for instance, Li et al. 2014). This might be due to the fact that, to perform a MT inference, participants do not only
need to process a conditional, but, additionally, a negation. This could also explain that AC is more frequently endorsed than DA.

Nevertheless, there are contexts in which MT does not seem to be applicable. To begin with, MT applied to conditional sentences whose consequents involve a deontically interpreted modal auxiliary verb like *should* or *must*, of which we can see examples in (30), results in arguments that are, to say the least, rather awkward:

(34) a. If you want to become a professional cellist, you must practice regularly.
   It is not the case that you must practice regularly.
   ? *Therefore*, you do not want to become a professional cellist.

b. If I don’t want to overpay, I should book my flight in advance.
   It is not the case that I should book my flight in advance.
   ? *Therefore*, I want to overpay.

c. If you are interested in conditionals, you should read Jonathan Bennett’s book.
   It is not the case that you should read Jonathan Bennett’s book.
   ? *Therefore*, you are not interested in conditionals.

One way to escape the problems with MT applied to deontic conditionals is by arguing, for instance, that their logical form, as opposed to the surface structure, is somehow different from simple $\varphi \supset \psi$. However, the material interpretation of a conditional does not allow us to make a distinction between sentences with and without modal auxiliaries. MT is a valid argument scheme and it should be applicable to any sentences that fall under the scheme without exceptions. Sentences in (34) definitely fall under that scheme.

In order to avoid the above discussed problems with deontic conditionals, advocates of the material interpretation can opt for excluding statements involving modal verbs from the analysis. They could claim that what they propose is a semantics for a fragment of natural language that consists of simple, atomic sentences and sentences that can be build thereof by means of logical connectives: $\neg$, $\lor$, $\land$, and $\supset$. However, this solution does not only appear to be rather *ad hoc*, but it also fails to settle all the issues related to MT and the material interpretation of the conditional. Consider the following instances of MT inferences (cf. Adams 1988):
(35)  a. If Dora dyed her hair, she didn’t dye it blue.
    Dora dyed her hair blue.
    Therefore, Dora didn’t dye her hair.

    b. If Eric bought a computer, he didn’t buy a Mac.
    Eric bought a Mac.
    Therefore, Eric didn’t buy a computer.

    c. If Patrick is running, he is not running fast.
    Patrick is running fast.
    Therefore, Patrick is not running.

We can easily imagine contexts in which the conditionals in (35) are fully assertable. Yet it would be unreasonable to allow the MT inference in these and similar cases. MT is classically valid and thus it must always be applicable to instances of material conditional. After all, $\phi \supset \psi$ is, by the law of contraposition, logically equivalent to $\neg\psi \supset \neg\phi$. MP and $\neg\psi$ suffice then to conclude $\neg\phi$. The following pairs of natural language conditionals and their contrapositives clearly show that those conditionals cannot be interpreted as material:

(36)  a. If Dora dyed her hair, she didn’t dye her hair blue.
    ? If Dora dyed her hair blue, she didn’t dye her hair.

    b. If Eric bought a computer, he didn’t buy a Mac.
    ? If Eric bought a Mac, he didn’t buy a computer.

    c. If Patrick is running, he is not running fast.
    ? If Patrick is running fast, he is not running.

Speaking colloquially, the contrapositives do not make sense at all. A material interpretation of a conditional allows contraposition because it ignores any possible relations or dependencies between a conditional antecedent and its consequent. In the above examples, the antecedents of the contrapositives entail the negations of their consequents. That Dora dyed her hair blue entails that she dyed her hair. As Mac is a brand of computers, buying a Mac entails buying a computer. And, obviously, Patrick’s running fast presupposes his running in the first place. The very possibility, not to even mention their prevalence, of this kind of analytic dependencies between constituents of a conditional undermines MT as a generally valid inference form for ordinary language conditionals. This is the first of the long list of difficulties into which the advocates of the material interpretation of a conditional are bound to run.
One could object that the above conditionals belong to a class of so called *non-interference conditionals*, that is sentences whose subordinate clauses can be introduced by means of “even if:”

(37) a. Even if Dora dyed her hair, she didn’t dye her hair blue.
    
    b. Even if Eric bought a computer, he didn’t buy a Mac.
    
    c. Even if Patrick is running, he is not running fast.

As such, they may need to be treated separately, analogously to the class of speech-act conditionals briefly discussed in the previous section. It is not clear, however, that the material account allows for differentiating between types of conditional sentences. Moreover, once again, this does not resolve the problem since there are cases of conditionals that do not fall into any special class (at least not in a sense that would be appropriate from the perspective of the advocates of the material account), yet their contrapositives are infelicitous, for instance:

(38) a. If Martha has not received any formal education, she is very talented.
    
    b. If Martha is not very talented, she has received a formal education.

MT inference does not seem to be applicable here either. Suppose that I believe that “If Martha has not received any formal education, she is very talented” is true. Upon learning from her teacher that Martha is not very talented, I might be more inclined to withdraw my believe in a conditional, or at least to lower my confidence that there is a meaningful connection between that conditional’s antecedent and its consequent, than to conclude that the girl has not received any formal education.

Perhaps even more striking are those cases in which Modus Ponens seems to fail. One of the most famous counterarguments against MP comes from McGee (1985). Before the 1980 elections, one had good reasons to believe that:

(39) If a Republican wins the election, then if it’s not Reagan who wins it will be Anderson.

given that there were two Republican candidates, Ronald Reagan and John Anderson. The opinion polls showed that Reagan was significantly ahead of the second in the race, the Democrat Jimmy Carter, with Anderson being a distant third, justifying a belief that:
(40) A Republican will win the election. However, as McGee observed, it would not be rational to believe the following conditional:

(41) If it’s not Reagan who wins, it will be Anderson.

even though one can arrive at it by virtue of MP. This example shows, then, that MP is not strictly valid. Moreover, experimental data on reasoning with right-nested conditionals suggest that MP inferences can be strong or weak depending on context (Huitink 2012; we will return to this issue at the end of chapter 3).

One could try to defend the validity of Modus Ponens by not allowing nesting of conditionals. After all, as has been already mentioned, MP seems to be one of the easiest inference patterns, and people strongly tend to endorse it. However, in an influential reasoning experiment, Ruth Byrne (1989, 1991) showed that a logically valid inference can be suppressed when an additional piece of information is added to the context. She reports that even though 96% of participants endorse the following valid inference:

(42) If she meets her friend, then she will go to a play.
    She meets her friend.
    Therefore, she will go to a play.

only 38% do so when a second conditional premise is added:

(43) If she meets her friend then she will go to a play.
    If she has enough money then she will go to a play.
    She meets her friend.
    Therefore, she will go to a play.

In classical logic and, consequently, on the material account of a conditional, the above argument is still valid, and hence people who do not endorse its conclusion commit a fallacy. The rules of classical logic notwithstanding, it often seems completely rational for people to withdraw earlier endorsed conclusions upon learning a new piece of information. Even though the following inferences:

(44) a. If Bob exercises twice a week, he will maintain his weight.
    Bob exercises twice a week.
    Therefore, Bob will maintain his weight.

b. If the switch is on, the lamp is on.
    The switch is on.
    Therefore, the lamp is on.
c. If Molly got an A for the logic course, her parents are proud of her.
Molly got an A for the logic course.
Therefore, Molly’s parents are proud of her.

are all instances of MP, and thus both valid and intuitively appealing, it suffices to add an additional premise to make their conclusions difficult to maintain:

(45) a. If Bob exercises twice a week, he will maintain his weight.
Bob exercises twice a week.
Bob eats only fast foods and drinks only sweetened sodas.
? Therefore, Bob will maintain his weight.

b. If the switch is on, the lamp is on.
The switch is on.
There is no light bulb in the lamp.
? Therefore, the lamp is on.

c. If Molly got an A for the logic course, her parents are proud of her.
Molly got an A for the logic course.
Molly failed all of her other exams.
? Therefore, Molly’s parents are proud of her.

Insisting that, for instance, Molly’s parents are still proud of her despite the fact that she failed everything but logic seems irrational, and so is holding on to the conclusion that the lamp without a light bulb is on just because the switch is on, or that Bob will maintain his weight regardless his unhealthy diet. Arguments in (44) and (45) are perhaps even more striking than those used by Byrne in her experiments, because the additional premise is here in a rather overt conflict with the conditional, and yet the argument is still logically valid, whereas in (43) the additional conditional premise only triggers an inference that makes the primary premise insufficient for the conclusion.

The failure of MP as illustrated by the above examples indicates that, unlike classical logic inferences, our everyday conditional reasoning is inherently defeasible. In addition, various studies on the role of background knowledge in conditional reasoning seem to indicate that it is highly context-sensitive (see, for instance, Thompson 1994; Liu 2003 or Klauer et al. 2010 on the effect of the perceived necessity and sufficiency of a conditional’s antecedent for its consequent on the evaluation of MP, DA, AC
and MT; or Thompson and Evans 2012 on so-called belief bias). Treating natural language conditionals as material implications is hence highly problematic. To make this point even stronger, let us consider a number of phenomena known as the paradoxes of material implication.

In classical logic and, consequently, under the material account of a conditional, strengthening of the antecedent is a valid argument form:

$$\varphi \supset \psi$$

$$\frac{(\varphi \land \chi) \supset \psi}{\varphi \supset \psi}$$

This rule leads to the first of the paradoxes of material implication, closely related to suppression of MP inferences demonstrated in (43) and (45). Consider the following pair of conditionals:

(46) a. If Molly got an A for the logic course, her parents are proud of her.

b. If Molly got an A for the logic course and failed all the other exams, her parents are proud of her.

Anyone who believes the first conditional is automatically committed to accepting\(^7\) the second even though the piece of information added to the antecedent of that conditional makes its consequent less likely to be true, and hence the whole sentence is hardly acceptable. Importantly, this is not an isolated case as the following pairs of conditionals clearly demonstrate:

(47) a. If Bob exercises twice a week, he will maintain his weight.

b. If Bob exercises twice a week and eats only fast food, he will maintain his weight.

(48) a. If you offer John a cup of tea, he will be pleased.

b. If you offer John a cup of tea and add a tablespoon of salt to it, he will be pleased.

(49) a. If the weather tomorrow is nice, I will go for a bike ride.

b. If the weather tomorrow is nice and I break my leg today, I will go for a bike ride.

---

7 I do not assume any theory of acceptability. The term is used in its intuitive, ordinary sense.
In each of the above pairs, the material account rules that we cannot believe the first conditional without believing the second, even though they sound absurd.

The paradoxes of material implication are even more taxing if they do not involve any changes in the contexts or in the belief states of a speaker. As a matter of fact, the material account allows us to generate countless instances of true yet absurd conditionals precisely due to the way their truth conditions are specified. The first class of problems stems from the fact that a true consequent is a sufficient condition for a materially interpreted conditional to be true, that is:

\[ \begin{align*}
\psi \\
\varphi \supset \psi
\end{align*} \]

is a valid argument. Therefore, if I believe the following sentences to be true:

(50) a. Nanga Parbat has never been climbed in winter.
    b. I have a younger sister.
    c. Lithuania is not a monarchy.

the material account commits me to accept the following conditionals, too:

(51) a. If summits of all 14 eight-thousanders have been reached in winter, then Nanga Parbat has never been climbed in winter.
    b. If my only sister is 5 years older than me, then I have a younger sister.
    c. If Lithuania is reigned by a king, then it is not a monarchy.

The second class of paradoxical conditionals owes its problematic character to the fact that a false antecedent is, again, sufficient for the truth of a material conditional. Given that

\[ \begin{align*}
\neg \varphi \\
\varphi \supset \psi
\end{align*} \]

is a valid argument form, the following conditionals:

(52) a. If aubergine is a species of small birds, then most Belgians speak Basque.
b. If Orhan Pamuk did not win the Nobel Prize in literature, then he is not a writer.

c. If raccoons are not American mammals noted for their intelligence, then they are not animals.

if interpreted materially, must be evaluated as true when the negations of their antecedents are evaluated as true:

(53)  
   a. Aubergine is not a species of small birds.
   b. Orhan Pamuk won the Nobel Prize in literature.
   c. Raccoons are American mammals noted for their intelligence.

The sentences in (51) and (52) seem so awkward, that one could think that no theorist could seriously defend the material account of a conditional. However, some philosophers, and most prominently Grice (1989), argued that there is nothing wrong with the above sentences in terms of their truth values. They all can be true yet simply unassertable. According to Grice, asserting a conditional when one knows the truth value of any of its constituents is a violation of one of the principles of good conversation, namely:

**THE PRINCIPLE OF QUANTITY:** Make your contribution as informative as is required (for the current purposes of the exchange).

This Gricean principle teaches us that a speaker who knows \( \varphi \) to be true but asserts a disjunction \( \varphi \lor \psi \) is highly uncooperative, for he does assert something weaker than what he has evidence for. Consequently, since the material conditional, \( \varphi \supset \psi \), is defined as equivalent to the disjunction of \( \neg \varphi \) and \( \psi \), asserting (52a) by a speaker who knows that aubergine is not a bird is as infelicitous as asserting:

(54) Either aubergine is not a species of small birds or most Belgians speak Basque.

in the same context. The same holds for asserting “Either Orhan Pamuk won the Noble Prize in Literature or he is not a writer” or “Either raccoons are American mammals noted for their intelligence or they are not animals” when one knows that, respectively, (53b) or (53c) are true. There is nothing wrong though with judging these disjunctions true, and, as Grice argues, neither there
is anything wrong with judging the corresponding conditionals true for they are unassertable for purely pragmatic reasons.

A Gricean defence of material conditional account has been motivated by semantic Occamism which teaches us not to multiply senses beyond necessity. Grice worried that interpreting “or,” “and,” and “if... then...,” as meaning something more than the logical connectives $\lor$, $\land$, and $\supset$, makes the connectives ambiguous, because there will always be a context in which some of the senses will be missing (see Bennett 2003, pp. 25-28, for a discussion of semantic Occamism). He prefers to explain the differences in how people use certain words as pragmatic phenomena. However, this strategy leads to a result that seems to betray Occam’s spirit itself, namely, to an unlimited multiplication of truths. That one should apply Occam’s razor to multiple meanings of a connective rather than to the profusion of nonsensical yet true propositions like (51) and (52) appears to be an arbitrary decision.

Let us assume, nevertheless, that allowing such a plethora of silly sentences to be true is not something to be worried about. As Edgington (1995, p. 243) provocatively proposes:

> If a theory which serves us well most of the time has the consequence that all such uninteresting conditionals are true, perhaps we can and should live with that consequence. It is too much—or maybe too little—to expect our theories to match ordinary usage perfectly. Perhaps, in the interests of simplicity and clarity, we should replace “if” with “$\supset$.”

But is it indeed the case that the material interpretation of “If $\varphi$, then $\psi$” aided by Gricean principles of good conversation suffices to explain all the data?

In fact, Edgington’s own response to the above suggestion is negative. As she observes, when we have to deal with beliefs that are not certain—which is arguably what natural language speakers usually do—the unacceptability of the inference from $\neg \varphi$ to $\varphi \supset \psi$ is even more striking. Edgington notes that when one has good reasons to believe $\neg \varphi$ yet they have no absolute certainty that it is true, they are justified in believing $\varphi \supset \psi$. For instance, to paraphrase Edgington’s own example, I am pretty confident that the king of the Netherlands is not on a visit to Poland right now. I am following the news and I am convinced that if such a visit were taking place, Polish newspapers would write about it. However, there is still a small chance that the Dutch king is actually in Poland but I have simply missed the news, or, perhaps
an even smaller chance, that he is now visiting the country incognito. Yet my high certainty that the king of the Netherlands is not in Poland at the moment leads me to accepting the following conditional:

(55) If the king of the Netherlands is in Poland right now, he is thinking about me.

assuming that the conditional is to be interpreted materially. This obviously is an absurd thing to believe or to assert. Nonetheless, uncertainty about $\neg \varphi$ means that $\varphi$ is not entirely excluded, and hence one cannot simply apply Gricean’s principles of good conversation to dismiss this and similar conditionals as true but unassertable.

The above conditionals with uncertain antecedents are not the only problems of the material interpretation that cannot be accounted for pragmatically. As Frank Jackson (1979) points out (I follow Bennett 2003, p. 32, here), logically equivalent sentences can differ in assertability, but Gricean pragmatics cannot explain such differences. For instance, the material account rules that as $\neg \varphi \land (\varphi \supset \psi)$ and $\neg \varphi \land (\varphi \supset \chi)$ are logically equivalent, because they are both equivalent to $\neg \varphi$. Hence, there should be no difference in assertability of the following sentences:

(56) a. The sun will come up tomorrow, but if it doesn’t it won’t matter.

b. The sun will come up tomorrow, but if it doesn’t it will be the end of the world.

Grice’s “assert the stronger instead of weaker” is of no help here as the above sentences are equally strong. Yet it does not only seem rational to assert just one of them and reject the other, but it could be easily considered irrational to even accept both of them. Intuitively, they appear contradictory.

Analogously, Gricean pragmatics does not help to explain the discrepancies between the contrapositives. As has been already mentioned, the contrapositives of the conditionals listed in (36) or (38). To repeat an earlier example, out of the following pair of conditionals:

(38) a. If Martha has not received any formal education, she is very talented.

b. If Martha is not very talented, she has received a formal education.
I may be inclined to assert the first but not the second, and there does not seem to be anything irrational about my preferences. Again, none of Grice’s principles of good conversation can explain why one of the two equivalent sentences is more assertable than the other. Yet again, material implication proves not to be the right interpretation of a natural language conditional.

Furthermore, interpreting conditionals materially and discarding all their odd instances on pragmatic grounds can have severe consequences for our epistemic hygienics. As has been discussed earlier, sentences like (52b) “If Orhan Pamuk did not win the Nobel Prize in literature, he is not a writer” are not assertable for someone who knows that the antecedent is false. Unassertability does not prevent us though from believing that if Orhan Pamuk did not win the Nobel Prize in literature, he is not a writer. After all, it is true that either Orhan Pamuk won the Nobel Prize in literature or he is not a writer, because he has actually been awarded the Nobel Prize in Literature. If such a belief is stored in the form of the above conditional, however, it may lead to certain false convictions.8

Imagine, for instance, that Bob, who believes (52b), encounters the name of Harry Mulisch, one of the most important Dutch writers of the last century. Mulisch, however, has never been honoured by the Swedish Academy, and therefore Bob may come to believe that he cannot be a writer at all. After all, (52b) that he already believes suggests that being awarded the Nobel Prize is some sort of a condition that has to be fulfilled for someone to be called a writer. It relates to the fact that (52b) could be paraphrased by Bob as “If someone did not win the Nobel Prize in Literature it means that he is not a writer” or “only people who win the Nobel Prize in Literature are writers.” Regardless whether these sentences can indeed be taken as correct paraphrases of (52b), they seem to be likely interpretations of the original sentence.

The reason for the misleading effect of a belief stored in a form that would make it unassertable, and in particular, as in this case, of a conditional whose truth is granted by the falsehood of its antecedent (or, analogously, by the truth of its consequent), is that such a conditional sentence conveys the existence of a connection between its antecedent and the consequent. A connection or some sort of a dependency between the clauses seems to be what we learn when we learn a conditional. If the material ac-

count deserves the title of the least intuitive theory of conditionals, it is precisely because it utterly ignores any semantic relations between the contents of their constituents. In its pursuit of truth-functionality, the material conditional account looses what is the most distinctive about conditional constructions, namely, their conditionality.

2.3 in search of conditionality

The counterintuitive consequences of the material account suggest that the natural language conditional requires stronger than truth-functional truth conditions. The data on how people actually use and interpret conditional sentences suggest that an empirically adequate theory should be guided by an idea that the antecedent of a conditional is some sort of a condition that the consequent in some way depends on.

Some of the most prominent propositional theories of conditionals that take the conditionality of conditionals into consideration have been inspired by a celebrated idea of Frank Ramsey (1990, p. 155), mentioned in section 1.2. Ramsey outlined a procedure for fixing one’s degree of belief in a conditional, known henceforth as the Ramsey Test, which can be construed as a specification of the acceptability conditions for conditional sentences. He argues that deciding whether to accept “If \( \varphi \), then \( \psi \)” or “If \( \varphi \), then \( \neg \psi \)” amounts to deciding whether \( \psi \) holds under the supposition that \( \varphi \). The test seems to hint at a plausible cognitive mechanism underlying the interpretation of conditional sentences, sometimes referred to as the simulation heuristics (Kahneman and Tversky 1982) or, more broadly, the process of hypothetical thinking (Evans and Over 2004, p. 153).

Being merely a footnote, however, a side comment that the author did not elaborate on, Ramsey’s illustrious passage allows some freedom of interpretation. Not surprisingly then, it inspired a family of theories focusing on conditional probabilities and acceptability conditions for conditionals, sometimes denying that conditionals can be truth-apt at all (Adams 1965; Edgington 1995). But it also did not stop Robert Stalnaker from incorporating the

---

10 It should be pointed out that applications of this mechanism go beyond the interpretation and production of conditional sentences. It has been proposed, for instance, that a process similar to the Ramsey Test is involved in something prima facie very different from evaluating conditionals, namely, in reasoning about other people’s beliefs (Peterson and Riggs 1999; Krzyżanowska 2013).
Ramsey test into his propositional account of conditional sentences.

Stalnaker (1968) developed Ramsey’s idea into a full-fledged truth-conditional semantics for both indicative and subjunctive conditionals. It is worth noting that some philosophers, and most notably Lewis (1973), are of the opinion that Stalnaker’s theory provides roughly the right framework for analysing subjunctive or counterfactual conditionals, and not indicatives. Lewis himself advertised the material interpretation of indicative conditionals which, as I have argued in the previous section, is not a descriptively correct account of natural language conditionals. Moreover, there seems to be a consensus in linguistics that semantics of indicatives is intensional (see, e.g., Kratzer 1986, 1991; Veltman 1985; von Fintel 2011). In fact, a view that became a dominant paradigm in linguistic literature, that is, so-called restrictor analysis developed by (Kratzer 1986) can be seen as a version of Stalnaker–Lewis approach.\footnote{As noted by von Fintel (2011), the restrictor approach is not really an alternative to the Stalnaker–Lewis semantics, “but a radical rethinking of the compositional structure of conditional sentences.”} Given that this thesis is predominantly concerned with indicative conditionals, I will focus on evaluating Stalnaker’s theory as an attempt to provide a semantics for conditionals in this class.

Stalnaker’s account is built upon the possible worlds semantics proposed by Kripke (1963). Note that even though the notion of a possible world seems metaphysically laden, Stalnaker emphasises that it is just a formal tool, a utility helpful in unravelling the structure of thought.\footnote{The notion of a possible world has also been recognised as a handy tool in psychology, where it helps to explicate people’s hypothetical and counterfactual reasoning, that is their ability to think about or to mentally represent alternatives to reality (Rafetseder et al. 2010; Leahy et al. 2014).}

Possible worlds are primitive notions of the theory, not because of their ontological status, but because it is useful to theorize at a certain level of abstraction, a level that brings out what is common in a certain range of otherwise diverse activities. The concept of possible worlds that I am defending is not a metaphysical conception, although one application of the notion is to provide a framework for metaphysical theorizing. The concept is a formal or functional notion, like the notion of an individual presupposed by the semantics for extensional quantification theory.\ldots
...The theory leaves the nature of possible worlds as open as extensional semantics leaves the nature of individuals. A possible world is what truth is relative to, what people distinguish between in their rational activities. To believe in possible worlds is to believe only that those activities have a certain structure, the structure which possible worlds theory helps to bring out (Stalnaker 1984, p. 57).

Possible worlds should be then understood as ways the world might be or might have been, or in other words, as alternative scenarios according to which things might happen or might have happened. In this sense, if we are about to toss a coin and wonder how it will land, we are considering two possible worlds, or, to be more specific, sets of worlds: one in which the coin lands heads, and another one in which it lands tails. If the coin has already been tossed and it landed heads, we may still entertain a thought that it might have landed tails. The notion of a possible world is, at least in this framework, just a useful theoretical device allowing us to model phenomena that involve thinking about various possibilities. For this reason, a possible worlds semantics seems perfectly well suited for analysing the meaning of conditional sentences.

Intuitively, the if-clause of a conditional invites us to imagine that the world is somewhat different, that is, it is different in a way specified by the content of the if-clause. The main clause can be then construed as an assertion that should be evaluated in relation to that mental image. For instance, when I am asserting or thinking that:

\[(57) \text{If Tolstoy did not write } \textit{Anna Karenina}, \text{ someone else must have.}\]

I first imagine that, contrary to my belief otherwise, \textit{Anna Karenina} was not authored by Tolstoy, and, consequently, in order to make sense of this imagined state of the world, I decide that the book must have been written by some other author. For this reason, I believe that \((57)\) is true. By contrast, I would not assert that:

\[(58) \text{If Tolstoy did not write } \textit{Anna Karenina}, \text{ the book does not exist.}\]

13 Some philosophers, and most notably Lewis (1973), disagree that the notion of a possible world can be treated instrumentally, as a metaphysically innocent tool.
even though it is easy to conceive of a possible world in which there is no such a book. The clear intuitions I have regarding (57) and (58) suggest that we intuitively impose certain constraints on what we can take as an alternative to the reality when asserting or interpreting conditional sentences.

To account for the difference between sentences like (57) and (58), a possible world semantics for conditional sentences must make it possible to determine which scenarios are, roughly speaking, “more possible” than others. In the Stalnaker’s framework, this is achieved by means of a selection function $s$ (or $s$-function for short) that picks the closest possible world in which the antecedent of a given conditional is true. More precisely, $s$ is defined as the mapping:

$$s : W \times 2^W \rightarrow W$$

(2)

where $W$ is the set of all possible worlds. The first argument of the selection function, $w_0 \in W$ is the candidate for the actual world, or, more generally, a world in which a conditional $\varphi \succ \psi$ is to be evaluated, and the second argument is the proposition expressed by the $\varphi$, that is, a set $\{w \in W : v(\varphi, w) = 1\}$ where $v : W \rightarrow [0,1]$ is a valuation function assigning binary truth values to propositions. The output of $s$ is a possible world $w_1 \in W$ such that, of the worlds in which $\varphi$ is true, $w_1$ is the most similar to $w_0$. To evaluate a conditional “$\varphi \succ \psi$”, one has to evaluate its consequent, $\psi$, in $s(\varphi, w)$, that is the closest $\varphi$-world. A conditional is true if and only if its consequent is true in the closest possible world in which the antecedent is true. More formally:

$$v(\varphi \succ \psi, w) = 1 \text{ if and only if } v(\psi, s(w, \varphi)) = 1$$

(3)

Assuming that a person’s beliefs, as propositions, can be represented as sets of possible worlds, the notion of the closest possible world neatly corresponds to Ramsey’s idea of minimally changed stock of beliefs. What an agent believes determines which possible worlds they consider as the candidates for the actual world. Upon hypothetically accepting a new proposition, an agent considers an alternative to the actual world, or rather, to what they believe is the actual world. The selected world $w'$ differs from the (candidate for the) actual one only with respect to the truth value

---

14 We are following here a convention of using the symbol “$>$” as a truth-conditional connective for Stalnaker’s nearness conditionals.
of \( \varphi \) and, if necessary, the values of other propositions that need to be changed in order to avoid inconsistency.

Closeness is a vague notion and it is somewhat controversial whether it can serve the purpose it has been designed for. Nevertheless, studies on counterfactual reasoning, and in particular, on so called counterfactual emotions of regret and relief suggest that people tend to have rather clear intuitions about the relative closeness of different possibilities (see, for instance, Kahneman and Miller 1986; Byrne 2002; Roese 2004; Teigen 2005). The examples of (57) and (58) well illustrate that our intuitions about plausibility of different scenarios can be fairly strong. I can easily imagine that Tolstoy did not write *Anna Karenina*. However, I am convinced that there is such a novel: I do not only remember reading it but I also own a copy. If I hypothetically accept “Tolstoy did not write *Anna Karenina*,” that is, if I imagine a possible world in which Tolstoy did not write this book, I do not have to hypothetically withdraw my belief that it has been actually written. Quite the contrary: it would be a waste of mental energy to make so far-reaching changes in the set of my beliefs since the thought of *Anna Karenina* not being one of the Tolstoy’s masterpieces can be easily reconciled with the belief that the book nonetheless exists. The possible world in which someone else than Tolstoy wrote *Anna Karenina* seems more similar to what I believe to be the actual world than the world in which the book has not been written at all. This is why I am inclined to judge (57) as true and (58) as false.

The difference between the two sentences can be easily illustrated by means of the following model:

Let \( p \) stand for “Tolstoy wrote *Anna Karenina*” and \( q \) for “The book titled *Anna Karenina* exists.” In the actual world, \( w_0 \) both \( p \) and \( q \) are true. Given that \( p \) entails \( q \), there is no possible world in which \( \neg p \land \neg q \) is true.\(^{15}\) The worlds \( w_1 \) and \( w_2 \) are \( \neg p \)-worlds.

\(^{15}\) To be precise, in Stalnaker’s semantics, there is a notion of an absurd world, defined as a world that is not accessible from any possible world, and where, roughly speaking, everything holds. Stalnaker introduced this semantic device to account for the meaning of conditionals with impossible antecedents, e.g., “If circles were square, it would never snow in Groningen.” In his theory, all conditionals with impossible antecedents are trivially true, because everything is true.
Since the only difference between $w_0$ and $w_1$ is the truth value of $q$ while $w_0$ and $w_2$ diverge with respect to both $p$ and $q$, $w_1$ is evidently more similar to $w_0$ than $w_2$ is.

Note that although my belief in Tolstoy’s authorship is very strong, I can still consider it as a possibility that *Anna Karenina* was attributed to him by mistake. Therefore, the conditional I utter does not need to be a subjunctive. Interestingly, Stalnaker’s semantics was supposed to account for both indicative and subjunctive conditionals, that is, on his account they have the same truth conditions. However, indicatives are pragmatically constrained: the closest possible worlds in which the antecedent is true, $w_1$ in the above example, cannot have been ruled out by anything accepted in the context in which the conditional is being evaluated. Whenever the antecedent is known to be false, the conditional is a counterfactual and the subjunctive mode is required. But given that we are rarely absolutely certain about the truth or falsehood of contingent propositions, it seems possible to use the subjunctive form when the antecedent is just very unlikely, though not entirely ruled out. For this reason, I might find the indicative (57) true while rejecting the subjunctive:

(59) If Tolstoy had not written *Anna Karenina*, someone else would have done that.

At the same time, I might be perfectly inclined to evaluate:

(60) If Tolstoy had not written *Anna Karenina*, the book would not have existed.

as true. If we pair the indicatives (57) and (58) with their subjunctive counterparts, (59) and (60), it appears that to tell the indicatives and subjunctives apart we may need something more than a pragmatic constraint since it is possible for one person in one context to believe (57) and (60), or, though the result does not sound particularly smoothly, to even conjoin them in one sentence:

(61) If Tolstoy had not written *Anna Karenina*, the book would not have existed, but if he did not write *Anna Karenina*, someone else must have done that.

in the world in which these antecedents hold. As pointed out by Lewis (1973), however, “the absurd world is a technical convenience not to be taken seriously,” and Stalnaker’s semantics can do without it (p. 78). See Nute and Cross (2002) for an alternative take on impossible antecedents within Stalnaker’s framework.
The above examples suggest that when deciding on which world is the most similar to the actual one, we take different aspects of that world into account, depending on whether we assert an indicative or a subjunctive conditional. Hence, if both indicatives and subjunctives are to be modelled within the same possible worlds framework of Stalnaker, we will have to explain why the selection function picks different possible worlds depending on the mood of the conditional sentence.

The issue of distinguishing between indicative and subjunctive conditionals aside, the concept of the closest possible world seems perfectly suited for explicating our intuitions about sentences like (57) and (58). However, even it raises problems of its own. Needless to say, similarity is an extremely vague notion, and even if in some cases we can simply count the propositions whose truth values are different, this will not always suffice. For instance, we may want to evaluate \( \neg p > q \) in the world \( w_0 \) of the following model:

\[
\begin{array}{ccc}
w_2 & \longrightarrow & w_0 & \longrightarrow & w_1 \\
\neg p, \neg q, r & \quad & p, q, r & \quad & \neg p, q, \neg r
\end{array}
\]

Since \( p \) holds in our base world \( w_0 \), we need to select the closest possible world in which \( \neg p \) is true to see whether \( q \) holds in that world. But which of the \( \neg p \)-worlds are we supposed to select: \( w_1 \) or \( w_2 \)? The world \( w_1 \) differs from \( w_0 \) with respect to the truth values of two propositions, \( p \) and \( q \), but so does \( w_2 \). It differs from \( w_0 \) with respect to the truth values of \( p \) and \( r \). There does not seem to be any reasonable way to decide which of the two \( \neg p \)-worlds is the closest: \( w_1 \) is as similar to \( w_0 \) as \( w_2 \) is.

To illustrate that this is not just an artificial example that has nothing to do with natural language, consider the following scenario. Bonnie and Clyde went for a hike. While Bonnie was admiring the landscape, Clyde walked ahead and she lost sight of him. Bonnie followed Clyde and arrived at a crossroads. It is likely that Clyde just walked straight ahead, however, he could have also taken a turn. Bonnie considers two conditionals:

(62) a. If Clyde did not go straight ahead, he turned right.

b. If Clyde did not go straight ahead, he turned left.

As Bonnie has no evidence in favour of any of the two conditionals, it is unclear how she should evaluate them. Looking for the
closest possible world in which Clyde did not go straight ahead is a lost call here for it is impossible to decide, given all the data at hand, whether the “turned right”- or the “turned left”-world is the closest. This on its own, however, is not necessarily a problem for Stalnaker’s logic, even though it validates the principle of Conditional Excluded Middle (CEM):

\[(\varphi \succ \psi) \lor (\varphi \succ \neg \psi)\]  

(4)

As he argues in (Stalnaker 1981), the truth values of conditionals in pairs like (62) or like the following famous example of subjunctives from Quine (1959, p. 15):

(63) a. If Verdi and Bizet had been compatriots, Bizet would have been Italian.

b. If Verdi and Bizet had been compatriots, Verdi would have been French.

can be undetermined. Stalnaker acknowledges vagueness as an immanent element of natural language and deals with it by adopting van Fraassen’s idea of supervaluations (van Fraassen 1966). Nevertheless, the above example illustrates that the similarity between worlds is not such an easy notion to operate with as one would prima facie expect.

In fact, the idea of the closest possible world has been challenged in the literature by many authors (see Bennett 2003, ch. 11 and 12, for a review), usually in the context of counterfactual conditionals like:

(64) If the president had pressed that button, a nuclear war would have started.

whose consequent is supposed to be true in possible worlds that are dissimilar from the actual world in the extreme (Fine 1975). This problem, however, can be resolved by specifying what kinds of dissimilarity can and cannot be allowed. For instance, we should not allow “miracles,” that is, the laws of nature should not be violated (Lewis 1979). But our intuitions about the similarity of worlds can also be at odds with our evaluations of certain indicative conditionals, especially when they involve vague predicates like “tall” or “bald.” Consider the following context: Bruce has a head full of hair. However, while some of his male family members do not experience any hair loss whatsoever, others go completely bald later in their life. Assuming that baldness is, at
least to an extent, a genetic condition, we may come to believe that:

(65) If Bruce goes bald, he will be completely bald.

“Bald” is a paradigmatic example of a vague predicate, hence not only a person with no hair at all can be referred to as bald, but also someone with, for instance, 100 or 500 hairs. Hairs on Bruce’s head in the actual world count in thousands, but how many hairs does Bruce have in the closest possible world in which he is bald? If the similarity between the worlds is a decisive factor, any world in which Bruce is partially bald is closer to the actual world than a world in which he is completely bald. Yet, given the above described context, (65) seems perfectly justified.

But assume, for the discussion’s sake, that we can define the relation of similarity between worlds in a satisfactory way. Does then Stalnaker’s semantics provide us with the right framework for analysing indicative conditionals? Is it suitable for capturing our intuitions about conditionality? Indubitably, Stalnaker’s theory does a much better job than the material account in explaining the meaning of “if” and the relationship between a conditional’s antecedent and its consequent. However, as I am going to show, it is not good enough. For this reason, it is worthwhile to have a closer look at the differences and, perhaps most importantly, the similarities between the two accounts.

The possible worlds semantics for conditionals is clearly not a truth functional account, since knowing the truth value of $\phi$ and $\psi$ is not sufficient for determining the truth value of $\phi > \psi$:

<table>
<thead>
<tr>
<th>$\phi$</th>
<th>$\psi$</th>
<th>$\phi &gt; \psi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0 or 1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0 or 1</td>
</tr>
</tbody>
</table>

As the above table illustrates, conditionals with false antecedents do not have to be all true, because their evaluation does not hinge upon what is true or false in the actual world, at least not in a direct way. *Prima facie*, this seems to have allowed Stalnaker to escape those paradoxes of material implication that result from the falsehood of the antecedent, like, e.g., the sentences discussed above:
(52) a. If aubergine is a species of small birds, then most Belgians speak Basque.

b. If Orhan Pamuk did not win the Nobel Prize in literature, then he is not a writer.

c. If raccoons are not American mammals noted for their intelligence, then they are not animals.

Supposing that aubergines are birds should not lead us to consider a possible world in which Belgians speak Basque as such a world would be much more remote than a world in which Belgians speak Flemish or French, as they actually do. Similarly, given that Orhan Pamuk is actually a writer, in the closest possible world in which he did not win the Nobel Prize in literature, he still is a writer—otherwise that possible world would not be similar enough to the actual one, that is, not similar enough to be considered the closest. Finally, hypothetically changing a very specific characteristics of American mammals known as raccoons can be limited to supposing that, for instance, they are not as cognitively skilled as they are reported to be in the actual world. Hypothesising that they might be something else than animals would lead us far beyond the closest possible world. For these reasons, Stalnaker’s semantics renders all the sentences in (52) false.

By the same token, the truth of a consequent is not sufficient for the truth of a conditional. Even though it is true that the summit of Nanga Parbat has never been climbed in winter, that I have a younger sister, and that Lithuania is not a monarchy, the following sentences are false on Stalnaker’s account:

(51) a. If summits of all 14 eight-thousanders have been reached in winter, then Nanga Parbat has never been climbed in winter.

b. If my only sister is 5 years older than me, then I have a younger sister.

c. If Lithuania is reigned by a king, then it is not a monarchy.

It seems natural to imagine that, as Nanga Parbat is one of the 14 mountains higher than 8000 meters above see level, in the closest possible world in which all of these 14 mountains have been climbed in winter, Nanga Parbat’s summit has been climbed,
too. Likewise, in a possible world in which my only sister is 5 years older than me, “I have a younger sister” must be false or else it would be an inconsistent world. And given that, by definition, a country that has a monarch acting as a head of state is a monarchy, in the closest possible world in which Lithuania has a king, it should also be a monarchy.

Furthermore, in Stalnaker’s logic:

\[
\frac{\varphi > \psi}{\neg \psi > \neg \varphi}
\]

is not a valid principle, hence it does not render the pairs of sentences in (36) equivalent. For instance, even if we accept that if Dora dyed her hair, she didn’t dye it blue, we are not only free to reject the contraposition: “If Dora dyed her hair blue, she didn’t dye her hair,” but we are even obliged to do so, for in the closest possible world in which Dora dyed her hair blue, regardless of how far from reality such a world is, she dyed her hair in the first place. In the closest possible world in which Eric bought a Mac, Mac should still be a computer and thus “If Eric bought a Mac, he didn’t buy a computer” cannot be true even if “If Eric bought a computer, he didn’t buy a Mac” is. Similarly, running fast presupposes running, therefore “If Patrick is running, he is not running fast” does not entail that if Patrick is actually running fast, he is not running.

Strengthening of the antecedent is not a valid argument form in Stalnaker’s logic either. Therefore, someone who accepts that if Bob exercises twice a week, he will maintain his weight, is not immediately committed to accepting that if Bob exercises twice a week and eats only fast food, he will maintain his weight. It is likely that the closest possible world in which both “Bob exercises twice a week” (p) and “Bob eats only fast food” (q) is further away from the actual world (w0) than the world in which only the first conjunct is true, so it does not necessarily make “Bob will maintain his weight” (r) true:

\[
\begin{array}{ccc}
\text{w0} & \text{w1} & \text{w2} \\
\neg p, \neg q, \neg r & p, \neg q, r & p, q, \neg r \\
\end{array}
\]

This seems promising. The truth conditions proposed by Stalnaker make it seem as if there was a need for some sort of a dependency between a conditional’s antecedent and its consequent.
In his 1968 paper, Stalnaker notes that the connection between $\varphi$ and $\psi$ is indeed sometimes relevant for the evaluation of “If $\varphi$, then $\psi$”:

If you believe that a causal or logical connection exists, then you will add the consequent to your stock of beliefs along with the antecedent, since the rational man accepts the consequences of his beliefs (p. 101).

However, Stalnaker’s semantics does not always help to avoid the troubles that made us discard the material account as inadequate. Even though a false antecedent is not sufficient to make a conditional true, conditionals with false antecedents and true consequents will often turn out true despite the lack of connection between the clauses as is the case with the following sentences:

(66) a. If aubergine is a species of small birds, then raccoons are American mammals noted for their intelligence.

b. If it is never raining in Groningen, then Kazimierz Ajdukiewicz was a prominent Polish philosopher.

c. If Tolstoy did not write *Anna Karenina*, then Ljubljana is the capital of Slovenia.

The consequents of the three above sentences are all true in the actual world. Even if we hypothetically assume that aubergine is a species of small birds, we do not need to change our beliefs about raccoons. In fact, given that we should keep the change as minimal as possible, we must not change any beliefs that are not inconsistent with the assumed proposition. Therefore, in the closest possible world in which aubergine is a species of birds, raccoons are still American mammals noted for their intelligence. Analogously, the closest possible world in which it is never raining in Groningen should not differ from the actual one with respect to the status of Kazimierz Ajdukiewicz as a prominent Polish philosopher. Likewise, in the closest possible world in which Tolstoy did not write *Anna Karenina*, Ljubljana should still be a capital of Slovenia. All three sentences in (66) are rendered true by Stalnaker’s semantics:

if you already believe the consequent (and if you also believe it to be causally independent of the antecedent), then it will remain a part of your stock of beliefs when you add the antecedent, since the rational man does not change his beliefs without reason (*ibid.*).
Strangely enough, this is not only the case despite there being no connection between these conditionals’ antecedents and their consequents. As a matter of fact, this is only possible because the connection is missing, and hence hypothetically assuming the antecedent has no bearing whatsoever on the truth value of the consequent:

This phenomenon is related to another problem with Stalnaker’s possible worlds semantics for conditionals. Whenever an antecedent, $\varphi$, is true in $w$, the following holds:

$$s(w, \varphi) = w$$

In other words, if a conditional’s antecedent is true in the actual world, then the closest possible world needed for the evaluation of that conditional is the actual world itself. As a result, the truth conditions defined by Stalnaker coincide with the material account whenever the antecedent is true, which means that Stalnaker’s semantics validates the following principle:

$$(\varphi \land \psi) \vdash (\varphi > \psi)$$

Any two sentences that are true in the actual world conjoined by means of a connective “if” will make a true conditional. Therefore,

$$(67) \quad \text{a. If aubergine is a vegetable, then raccoons are American mammals noted for their intelligence.}$$

$$\text{b. If it sometimes rains in Groningen, then Kazimierz Ajdukiewicz was a prominent Polish philosopher.}$$

$$\text{c. If Shakespeare did not write } \text{Anna Karenina}, \text{ then Ljubljana is the capital of Slovenia.}$$

are all rendered true on Stalnaker’s account. Even though this theory appears to be all about a dependence between conditionals’ antecedents and their consequents, it still fails to pin down what this dependence exactly is. And given that it allows some of the undesirable consequences of the material account to sneak in, it also fails to adequately capture the meaning of a natural language conditional.

2.4 WHAT IS WRONG WITH THE RAMSEY TEST?

It seems that Stalnaker’s account fails exactly because of what the Ramsey Test does not take into account. The possible world semantics as well as the nonpropositional theories of conditionals
that have been build upon Ramsey’s idea (e.g. Adams 1965, 1975; Gibbard 1981; Edgington 1995) fail to acknowledge that there is more to the dependency between an acceptable conditional’s antecedent and its consequent than the truth or a high degree of belief in the consequent under the supposition of the antecedent. The Ramsey Test is perhaps best construed as an operational definition of conditional probability and as such it does not exhaust the meaning of conditional sentences, though it somewhat suggests a broader mechanism of hypothetical thinking involved in their production.

As a procedure of fixing one’s degrees of belief in a conditional, the Ramsey Test seems to be, intuitively, plausible. Therefore, it seems to present a good basis for the acceptability or assertability conditions for conditional sentences, and as such it has been utilised by many authors, with, perhaps, the most influential of them being Ernest Adams (1965, 1966, 1975) according to whom the acceptability or assertability of a conditional “goes by” its conditional probability. Regardless of how we specify what exactly “goes by” means, one is clear: “If \( \varphi, \psi \)” is highly assertable or acceptable if and only if \( \Pr(\psi | \varphi) \) is high. That there is a strong correlation between a conditional’s degree of acceptability or assertability and the probability of the conditional’s consequent on a supposition of its antecedent, usually referred to as The Adams Thesis, is, interestingly, one of the very few claims about conditionals that is not only relatively uncontroversial, but also frequently taken for granted, even though it has been experimentally shown that, as a general claim about how actual speakers use conditional sentences, it does not hold (Douven and Verbrugge 2010). Moreover, it is not unproblematic if taken solely as a normative principle. As observed by Douven (2008), the high conditional probability does not suffice for the high degree of assertability or acceptability of a conditional either. If it were sufficient, the following two conditionals would be equally assertable or acceptable:

\[
(68) \begin{align*}
\text{a. If there is a heads in the first ten tosses,} & \\
& \text{there will be at least one heads in the first 1,000,000} \\
& \text{tosses of this fair coin.}
\end{align*}
\]

\[
\begin{align*}
\text{b. If Chelsea wins the Champions League,} & \\
& \text{there will be at least one heads in the first 1,000,000} \\
& \text{tosses of this fair coin.}
\end{align*}
\]

See also Lewis (1973); Jackson (1979, 1987)
The probability of the shared consequent of the two above sentences, that is, the probability that there will be at least one heads in the first 1,000,000 tosses of a fair coin (ψ) is extremely high. More precisely, \( \Pr(\psi) = 1 - \frac{1}{2^{1,000,000}} \approx 1 \). On the assumption that there is a heads in the first ten tosses of that fair coin (α), ψ is absolutely certain, that is, \( \Pr(\psi | \alpha) = 1 \). Being a tautology, (68a) is perhaps not very likely to be asserted in a conversation, nevertheless, it is perfectly acceptable and assertable. (68b), by contrast, seems to have a rather low degree of assertability, yet it passes the Ramsey Test with flying colours: after hypothetically adding “Chelsea wins the Champions League” (β) to our stock of beliefs, we find out that our degree of belief in (ψ) is very high, thus the conditional (68b) should be highly acceptable and assertable. As long as \( \Pr(\beta) \) is non-zero, and \( \Pr(\psi) \approx 1 \), \( \Pr(\psi | \beta) \approx 1 \), thus, assuming the Adams Thesis, the degree of assertability or acceptability of a conditional with β as an antecedent and ψ as a consequent is extremely high. However, this leads to counterintuitive consequences, parallel to the paradoxes of the material implication. As the example of (68b) clearly shows, the Ramsey Test and the Adams Thesis allow for conditional sentences whose antecedents have nothing to do with their consequents to be rendered highly acceptable or highly assertable, which should not be the case.

The assertability or acceptability conditions defined on a basis of the Ramsey Test fall into a similar problem as Stalnaker’s semantics: they do not account for the connection between a conditional’s antecedent and its consequent in an appropriate way, forcing us to yield acceptable and assertable bizarre conditionals that, intuitively, are not acceptable nor assertable at all. But if the effect of assuming the antecedent on the probability of the consequent is not sufficient for the acceptability or assertability of conditional sentences, what is still missing? Within Bayesian framework, the hitherto missing connection between the antecedent and the consequent of a conditional can be captured by means of an evidential support relation (Douven 2008). Some logicians, on the other hand, talked about a relation of relevance, which according to some authors amounts to sharing variables, while others insisted that an antecedent of a conditional should be a part of a proof of that conditional’s consequent (Mares 2014).17 However, before we can choose between any of the available proposals, we should ask ourselves a question: what do we assert

17 The relevance logics will be briefly discussed in section 3.2 of the next chapter
when we assert a conditional sentence? In other words, what is it that we are trying to communicate whenever we decide to use a sentence of the form “If $\varphi$, $\psi$” instead of some simpler, more straightforward construction. In the following chapters, I will attempt to answer these questions and, accordingly, to propose a new semantics of conditionals that will do justice to both our intuitions and the linguistic data on these sentences are used and interpreted by natural language speakers.