

**You don’t say!**

**Lying, asserting and insincerity**

**By:**

Neri Marsili

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

The University of Sheffield

Faculty of Arts and Humanities

Department of Philosophy

July 2017

**Abstract**

This thesis addresses philosophical problems concerning improper assertions. The first part considers the issue of defining lying: here, against a standard view, I argue that a lie need not intend to deceive the hearer. I define lying as an insincere assertion, and then resort to speech act theory to develop a detailed account of what an assertion is, and what can make it insincere.

Even a sincere assertion, however, can be improper (e.g., it can be false, or unwarranted): in the second part of the thesis, I consider these kinds of impropriety. An influential hypothesis maintains that proper assertions must meet a precise epistemic standard, and several philosophers have tried to identify this standard. After reviewing some difficulties for this approach, I provide an innovative solution to some known puzzles concerning this problem. In my view, assertions purport to aim at truth, but they are not subject to a norm that requires speakers to assert a proposition only if it is true.

**Table of Contents**

[I. An introduction 11](#_Toc489295483)

[1. Lies 12](#_Toc489295484)

[2. Incorrect Assertions 16](#_Toc489295485)

[3. Outline of the Dissertation 19](#_Toc489295486)

[Part A- Lying 21](#_Toc489295487)

[A. The definition of lying 23](#_Toc489295488)

[1. The Classic Definition of Lying 23](#_Toc489295489)

[2. The Current Debate: Deceptionist vs Non-Deceptionist accounts 31](#_Toc489295490)

[II. Deceptionist definitions of lying: a critique 35](#_Toc489295491)

[1. The intention to deceive condition: an introduction 35](#_Toc489295492)

[2. Different versions of the intention to deceive condition 36](#_Toc489295493)

[2.1. The “broad” IDC 36](#_Toc489295494)

[2.2. The “rigid” IDC 37](#_Toc489295495)

[2.3. The “believed sincerity condition” and the disjunctive account 38](#_Toc489295496)

[3. Against sufficiency: non-assertoric falsehoods that intend to deceive 42](#_Toc489295497)

[4. Against necessity: lying without the intent to deceive 46](#_Toc489295498)

[4.1 Lying under duress 46](#_Toc489295499)

[4.2 Bald-faced lies! 48](#_Toc489295500)

[5. Deceptionist Replies 50](#_Toc489295501)

[5.1. Are bald-faced lies intended to deceive? 50](#_Toc489295502)

[5.2 Are bald-faced lies assertions? 53](#_Toc489295503)

[6. Conclusions 59](#_Toc489295504)

[III. Lying: a speech-act theoretic approach 62](#_Toc489295505)

[1. Assertion-based definitions, and their difficulties 63](#_Toc489295506)

[1.1. Carson 64](#_Toc489295507)

[1.2. Fallis 66](#_Toc489295508)

[1.3. Stokke 70](#_Toc489295509)

[2. Three Puzzles From Speech Act Theory 73](#_Toc489295510)

[2.1 Explicit Performatives 73](#_Toc489295511)

[2.2. Determining which speech acts can be lies 76](#_Toc489295512)

[2.3. Insincerity Conditions 81](#_Toc489295513)

[3. A Speech-act Theoretic Account of Lying 84](#_Toc489295514)

[3.1 The Assertion Condition 84](#_Toc489295515)

[3.2 Illocutionary Entailment 87](#_Toc489295516)

[4. Testing the definition 89](#_Toc489295517)

[4.1 Proviso Lies 89](#_Toc489295518)

[4.2 The Speech-act Theoretic Puzzles 91](#_Toc489295519)

[5. Conclusions 96](#_Toc489295520)

[IV. Insincerity 99](#_Toc489295521)

[1. Insincerity: a preliminary account 101](#_Toc489295522)

[2. Beyond belief: insincerity and illocutionary acts 103](#_Toc489295523)

[2.1 The third speech act theoretic puzzle 103](#_Toc489295524)

[2.2 Expressing attitudes and (in)sincerity 104](#_Toc489295525)

[2.3 Insincerity conditions for promising 106](#_Toc489295526)

[2.4 A general account of the insincerity conditions for lying 110](#_Toc489295527)

[3. Insincere Promises: an experimental study 112](#_Toc489295528)

[3.1 Testing folk intuitions about lying 112](#_Toc489295529)

[3.2 Aim of the study 113](#_Toc489295530)

[3.3 The predictions of existing theories 114](#_Toc489295531)

[3.4 Experiment 1 115](#_Toc489295532)

[3.5 Experiment 2 119](#_Toc489295533)

[3.6 Experiment 3 120](#_Toc489295534)

[3.7 General discussion 122](#_Toc489295535)

[3.8. More on the results 122](#_Toc489295536)

[3.9 Conclusions 127](#_Toc489295537)

[4. Gradedness 129](#_Toc489295538)

[4.1 The dichotomic view and the traditional insincerity condition 129](#_Toc489295539)

[4.2 Graded Truth Values 131](#_Toc489295540)

[4.3 Graded beliefs 134](#_Toc489295541)

[4.4 A “graded” definition of insincerity 135](#_Toc489295542)

[4.5 Expressing graded beliefs and graded truth values 140](#_Toc489295543)

[4.6 Further graded components 146](#_Toc489295544)

[5. Conclusions – A general account of insincerity 150](#_Toc489295545)

[Part B - The norms of assertion 155](#_Toc489295546)

[B. The norms of assertion 157](#_Toc489295547)

[V. A ‘constitutive’ norm? 161](#_Toc489295548)

[1. The orthodox account of constitutive norms 161](#_Toc489295549)

[2. Williamson on constitutive norms 163](#_Toc489295550)

[3. The C-Rule as a Constitutive Rule 165](#_Toc489295551)

[3.1. Phrasing the C-rule as orthodoxly constitutive 165](#_Toc489295552)

[3.2. An alternative constitutive reading 166](#_Toc489295553)

[3.3 Treating two rules as one 167](#_Toc489295554)

[4. The C-Rule as a Regulative Rule 169](#_Toc489295555)

[4.1 The C-rule as orthodoxly regulative 169](#_Toc489295556)

[4.2 Pollock’s paradigm: ‘prescriptive’ constitutive rules 170](#_Toc489295557)

[4.3 All norms are constitutive in Pollock’s sense 171](#_Toc489295558)

[4.4 The distinction is not meaningful 174](#_Toc489295559)

[5. Conclusions 175](#_Toc489295560)

[VI. Truth and assertion: rules vs aims 180](#_Toc489295561)

[1. The disagreement about the norm of assertion 180](#_Toc489295562)

[1.1 The norm of assertion 180](#_Toc489295563)

[1.2 Intuitions and norms 182](#_Toc489295564)

[1.3 Falsity-Criticism and inadvertent violations 183](#_Toc489295565)

[2. Lucky and unlucky assertions 185](#_Toc489295566)

[2.1 Lucky: inadvertent observation of TR 185](#_Toc489295567)

[2.2 Unlucky: inadvertent violation of TR/KR 186](#_Toc489295568)

[2.3 Checkmating the speaker: a too demanding norm 188](#_Toc489295569)

[3. Deriving permissibility from correctness 190](#_Toc489295570)

[4. Truth as aim 193](#_Toc489295571)

[4.1 The difference between rules and aims 194](#_Toc489295572)

[4.2. Truth as a rule, truth as an aim 195](#_Toc489295573)

[4.3 The truth-aim account 198](#_Toc489295574)

[4.4 Ought to try – from aims back to rules 200](#_Toc489295575)

[5. Objections, replies and clarifications 203](#_Toc489295576)

[5.1 The source of normativity 203](#_Toc489295577)

[5.2 Primary and secondary violations 204](#_Toc489295578)

[5.3 Challenges 209](#_Toc489295579)

[5.4 Knowledge as the aim? 210](#_Toc489295580)

[5.5 Alleged asymmetries 211](#_Toc489295581)

[6. Conclusions 212](#_Toc489295582)

[VII. Conclusions 214](#_Toc489295583)

[References 232](#_Toc489295584)

Ad Angiolino, Basilio e Riccardo

# I. An introduction

It will be convenient to start from the beginning. *In principio erat Verbum, et Verbum erat apud Deum, et Deus erat Verbum* (Jhn, 1, 1-3): “In the beginning was the Word, and the Word was with God, and the Word was God”. If in the beginning of time language was one with God and belonged to God, who is the source of all Truth (Pr, 8,7; Sam 7, 28; Rm 3, 4), its evil relative, the lie, appeared short after creation.

According to the Biblical myth, God told Adam: “Of every tree of the garden thou mayest freely eat: But of the tree of the knowledge of good and evil, thou shalt not eat of it: for in the day that thou eatest thereof thou shalt surely die” (Gen, 2, 16-17). But Satan, “the father of lies”, who “speaks his native language [when he lies]” (Jhn, 8, 44), under the fake appearance of a snake, told Eve: “Ye shall not surely die: For God doth know that in the day ye eat thereof, then your eyes shall be opened, and ye shall be as gods, knowing good and evil” (Gen 3, 4-6). If we believe the Bible, this was the very first lie ever told; and it was this lie that determined the expulsion of humans from paradise, and the beginning of our history.

In the Bible, truth and truthfulness associated with God and the Good; falsity and untruthfulness with the Devil and Evil. The precept “thou shalt not bear false witness against thy neighbour” (do not lie) figures in the Tenth Commandments that God gives to Moses on mount Sinai. The Bible is one of the foundational books of Western culture: it is not a coincidence that truthful and untruthful communications are central themes in this text, given their importance for humanity, and many are the parables and allegories that involve them. In a less allegoric fashion, and with more modest intentions, this dissertation will present and analyse, from a philosophical point of view, the opposition between these two opposite communicative forces.

Communication is a fundamental ability of human beings – it is one of the abilities that makes us distinctively human. Unfortunately, the ability to communicate comes at a price: it makes it possible (and easy) for speakers to misrepresent what they believe, to communicate something false; in other words, it allows communicators to lie. This concept is beautifully illustrated in this famous quote from Umberto Eco (1975):

*[The study of communication] is in principle the [study of] everything which can be used in order to lie. If something cannot be used to tell a lie, conversely it cannot be used to tell the truth: it cannot in fact be used "to tell" at all.*

Any signal that can be used to tell the truth can be used to tell a falsehood: the ability to communicate sincerely is essentially entangled with the ability to communicate insincerely. To some, it may appear that language is for this reason an intrinsically unreliable source of knowledge. After all, lying is not the only way in which language can deceive us. Not only speakers can be insincere: they can also be mistaken about what they say, and communicate falsehoods even when they are speaking in good faith. And still, despite the possibility of lies and mistakes, we believe most of what other people say.

Rather than a matter of mere gullibility, the trust we place in each other’s reports is the result of our dependency on human communication for survival. Most of the knowledge we use to lead our daily lives comes from what others tell us. It is by trusting teachers and authors of books that we apprehended everything that we learnt in school: from physics to geography, from grammar to the whole history of the world. Even our own names are something we learn from hearsay, as none of us can recollect those initial days of our life in which we were given one. Without the trust we place in communication, we would be paralyzed as epistemic agents, and it would be almost impossible for us to lead a meaningful life. And still, this does not mean that other people’s testimony is always reliable: it simply means that we cannot afford but to trust most of what we are told.

Considering these issues, it becomes quite clear that understanding truthful and untruthful communication is of central importance, both for the study of language and that of our ordinary life. The aim of this dissertation is to explore untruthful communication from a philosophical point of view. The dissertation is divided in two parts; with some approximation, we could say that each part touches the two sides of this coin: the first part (A) deals with the deliberate communication of something false, the second part (B) with the accidental communication of something false. In what follows, I will clarify more in detail what this means, and to what extent this description is an approximation.

### 1. Lies

The first part of this dissertation (Part A) will cover philosophical problems related to lying, insincerity and deception. These phenomena are of fundamental importance in contemporary society, where communication plays an increasingly important role. To see this, it will be helpful to look at some recent examples. If we limit the attention to politics, the current political debate has seen an unprecedented rise in lying and other forms of deceptive communication. Before the Brexit referendum in June 2016, UK media have infamously reported numerous false or deceptive claims. Among these, the false assertion that the NHS was “nearly at breaking point” due to “a massive influx of EU immigrants”, and that “more than 700 offences are being committed by EU immigrants every week” [[1]](#footnote-1). The most blatant of these lies was the pledge to convert EU spending into NHS funding by £350m-a-week. Painted on the side of Vote Leave's big red bus, this pledge was hardly realisable, based on fraudulent data (net EU contribution was merely 160m-a-week at the time), and readily disowned by Brexiters within a few hours after the election was won[[2]](#footnote-2).

The same year, the US presidential elections saw a similar upsurge of lies in political debates and campaigns. For Donald Trump’s in particular, lying has been a fundamental campaigning strategy. Journalist Maria Konnikova went as far as claiming that “the sheer frequency, spontaneity and seeming irrelevance of his lies have no precedent”[[3]](#footnote-3). This is hardly an exaggeration: a study shows that (out of a given sample) an estimate of 70% of Trump’s statements during the presidential campaign were false, another reported that in a one-hour-long TV appearance Trump managed to utter an astounding total of 71 lies[[4]](#footnote-4). Trump’s attitude has not changed since: this 23rd of June, six months after the elections, his dedication to lying has prompted the *New York Times* to dedicate a full page of the newspaper to print every lie that Trump has publicly told since taking office[[5]](#footnote-5).

This apparent increase in political lying and deception has brought some commentators to question whether the UK referendum and the US elections were based on informed voting, an essential ingredient to a functioning democracy. Similarly, the legitimacy of the political decisions taken on the basis of such voting has been repeatedly put to question. More generally, these events have brought to the fore the importance of understanding lying and other form of deceptive communication.

If understanding lies is important, it is not because of politics alone. Lies have been at the centre of the public debate for a number of other reasons. Famous sportsmen have infamously lied about taking performance-enhancing drugs, leading many supporters to question how fair these competitions are: if we consider the case of road cycling alone, famous are the cases of Tour de France winners Marco Pantani and Lance Armstrong, the latter of which was eventually stripped of all his titles. The job of a scientist is the pursuit of truth, but even academics are susceptible to the temptations of lying. One famous case is that of physicist Jan Hendrik Schön. His purported discoveries about molecular semiconductors were so revolutionary that the scientific world was expecting him to win the next Nobel Prize in Physics – then his data turned out to be fraudulent, Schön was fired and his PhD revoked[[6]](#footnote-6). More generally, humans are prone to lie, way more than we are usually willing to admit to ourselves: studies show that people on average report to lie one or two times a day (DePaulo et al. 1996) – but this rate can go as high as a mean of *three lies* *a minute* when we consider interactions with strangers (Feldman, Forrest & Happ 2002).

In light of these observations, is not surprising that the study of lying, insincerity and deception is gaining centre stage in studies of linguistic communication: disciplines as diverse as sociology, forensics, psychology, and neuroscience have displayed an increasing interest in their analysis (Levine 2014). Philosophy certainly cannot aspire to develop lie-detection techniques, and much less a ‘cure’ for lying, but it can certainly aim to understand this phenomenon, bring conceptual clarity to its study, and open the way for further investigation. Such is the purpose of the first half of this dissertation.

To give an idea of what kinds of philosophical issues arise with regard to lying, it will be helpful to have a quick overview of the main philosophical debates concerning the issue. Perhaps the oldest recorded philosophical interest in lying can be traced back to the sixth century B.C.: it is in these times that Epimenides of Knossos developed the ‘liar paradox’, challenging logicians to determine the truth value of statements of the form “I always lie” for the centuries to come.

Throughout history, however, the most discussed issue in the philosophy of lying has almost certainly been that of the morality of lying. Questions like “is it ever morally permissible to lie – and if so, under which conditions?” have gripped philosophers throughout history: from Augustine to Aquinas, from Kant to Grotius, many established moral philosophers have tried to tackle these conundrums. A related philosophical issue is that of the permissibility of political lying. A tradition from Plato to Leo Strauss has defended the claim that lying for the common good is permissible and recommended for a virtuous leader; an opposing tradition disagrees – amongst other reasons, on the ground that political lying undermines democratic institutions (cf. Carson forthcoming). More recently, philosophical interest in lying has emerged in the contemporary debate in epistemology. Around the 1980s, philosophers got more and more interested in the transmission of knowledge via testimony – *i.e.,* knowledge that epistemic agents learn from words. Within this field of inquiry, lies are philosophically significant, as they represent a potential hindrance to the social process of knowledge transmission.

Each of these philosophical questions (except for the liar paradox) presuppose a clarification of the concept of lying: to be able to tackle them, we need first to determine what lying is – and this, in itself, is a philosophical question. Rather than delving into any specific debate on lying, thus, the first part of this dissertation will be devoted to respond to this foundational question: “what is lying?”. More specifically, it will attempt to identify necessary and sufficient conditions for a statement to count as a lie. This apparently easy task will turn out to be one of the most challenging for the philosophy of lying: it is not easy to identify a definition that is resistant to counterexamples; furthermore, an informative analysis of lying demands further analysis of a number of neighbouring concepts – such as insincerity, assertion and deception.

In attempting to define a concept, this dissertation places itself in the tradition of conceptual analysis. A central idea within this tradition is that we can learn something about a given concept by breaking it down into its essential components. In this sense, the desideratum of a theory of lying is to identify the necessary and sufficient conditions for something to be a lie.

From a methodological point of view, this means that a good analysis of lying should neither be too broad (it should not be subject to counterexample to the sufficiency of the analysis) nor too narrow (there should not be counterexamples to its necessity). The litmus test for a good definition here are intuitions: there is consensus that a good definition should reflect our intuitions about particular cases. There is disagreement as to the scope of ‘our’ here – some scholars think that a good definition should reflect laypeople’s intuitions, other that expert intuitions are more important. Ideally, a good definition should meet the intuitions of both groups.

As I proceed in my analysis, I will argue in favour of a number of claims about lies. Some will be claims about conceptual possibilities and unusual lies: that there can be lies that are not intended to deceive; that there can be lies that are true, that there can be lies that we believe to be true. Even though each of these claims may sound astounding, I will show that they are perfectly consistent with our intuitions about particular cases, and relevant to the understanding of what lying is. Some other claims will be about the inability of traditional definitions to explain some standard features of lying: for instance, standard definitions struggle to explain that one can lie by promising something, that all lies always involve the undertaking of some distinctive responsibility, and that lies can be more or less insincere, depending on our degree of confidence of our beliefs, and on the hedges and other linguistic devices that we use to express such believes. As I review and criticise the existing literature, I will develop my own definition of lying – hopefully, one that will be useful to solve other philosophical problems about lying, such as ethical and epistemological ones. For a more detailed breakdown of the topics that will be touched in this part of thesis, the reader can refer tothe *Plan of the Work* (I.3).

### 2. Incorrect Assertions

While the first part of the thesis deals with insincere communication, the second part deals with irresponsible communication: statements that are at fault not because they are insincere, but rather because they are false, or because they are not supported by adequate evidence. There are uncountable real-world examples of this sort of communicative improperness; to find one to which we are all acquainted, it will be helpful to consider some popular conspiracy theories. Even if we limit our attention to wildly implausible (yet popular) ones, the list of such theories is endless: one claims that the Earth is hollow and secret civilisations live inside it, in the hidden city of Agarthaa, illuminated by an internal sun; another that the Earth is flat, and that governments *create* false evidence to trick us into believing the opposite; others that our planet has been visited by aliens, and governments operate to *destroy* all signs of contact with extra-terrestrial life.

What is at issue here is not why conspiracy theorists believe in these theories, or what is wrong with such beliefs. It is rather the role that conspiracy theorists have in the propagation of these falsehoods – their role as communicative agents. Conspiracy theorists are not liars[[7]](#footnote-7): when they assert that the Earth is flat or hollow, they do it in good faith, sincerely. But even if their assertions are not lies, they are importantly at fault.

In the age of social media, this form of irresponsible communication has become even more problematic and apparent. On the one hand, social media make it easier for people to share false information, without verifying its veracity. On the other hand, social media allow this information, shared in good faith, to reach a wider audience – often with catastrophic effects. One infamous example is the epidemic sharing of the fake news that Pope Francis endorsed Donald Trump during the US electoral campaign. Initially posted by the satirical fake news website WTOE 5 News, the news was shared by roughly a million of Facebook users, giving the rumour a resonating voice: it was the most shared news posted during the elections, outperforming any real one at the moment. Considering a larger sample, a study shows that pro-Trump fake news stories received around 30 million shares on Facebook during the electoral campaign, and often more attention than real ones[[8]](#footnote-8). Overall, this kind of data strongly suggests that the irresponsible communicative behaviour of some voters (their sharing unconfirmed news coming from dubious sources) might have strongly influenced the outcome of the last US elections.

The communication of false information and unwarranted claims thus poses a serious threat to a citizen’s access to reliable, factual information – and to some extent to democracy itself, as long as we understand this institution to be reliant on informed voting. In some cases, it can even be a threat to people’s lives and wellbeing. Such in the case of the disinformation spread by the anti-vaccine movement, that has led to an upsurge in deaths by otherwise curable diseases, with an estimated death toll of about 9 thousand people since 2007[[9]](#footnote-9).

I have already pointed out that conspiracy theorists and anti-vaccine activists are not lying when they propagate their falsehoods (since they believe in them), and that nonetheless their communicative behaviour is at fault. If not lying, which kind of communicative fault is committed in these cases? The second part of the dissertation attempts to answer this question: it aims to understand, on a general level, what kind of communicative expectation is violated when people assert false or unwarranted propositions. This sort of question has gained centre stage in the philosophical debate in recent times. For reasons that I will discuss more in detail later, in the last 20 years philosophers have become more and more convinced that assertions (claims that something is true) are governed by a single rule – a rule whose violation can explain the distinctive wrongness involved in uninformed and false statements. Scholars working in this tradition tend to adopt the following hypothesis (or a hypothesis along the following lines):

Assertion is the only speech act governed by the following rule: you should not assert a given proposition unless condition C is satisfied

According to this hypothesis, there is a condition that must be satisfied for an assertion to be appropriate, and if we replace ‘C’ with that condition, we obtain the norm that governs all assertions. This hypothesis offers a simple model to explain what is wrong with assertions by uninformed speakers (like conspiracy theorists and anti-vaccine activists): these assertions are at fault not because they are lies, but because they violate the norm governing all assertions. Just like we expect people not to lie, we expect speakers to follow the norm; when they fail to do so, a violation occurs, making ground for reproach, criticism and blame.

Given that the hypothesis leaves condition C unspecified, a great deal of the explanation will turn on how one specifies condition C. To see a few candidates, let us consider a real life scenario. Suppose that, after reading about the fake news that the Pope endorsed Trump on the satirical news site WTOE 5 News, an American citizen (call him Ferdinand) tells another friend that the pope endorsed Trump as a presidential candidate. Now, we might have the intuition that Ferdinand’s assertion is wrong because it is not true, which gives us a first candidate for condition C: do not assert a proposition *unless it is true*. But some other people might have the intuition that Ferdinand’s assertion is wrong for different reasons. For instance, Ferdinand does not have the appropriate evidence to make such a claim: he failed to check whether the information came from a reliable source (as a matter of fact, it did not), which is especially pernicious given the implausibility of the proposition (there are many known ideological disagreements between the Pope and Donald Trump). If one has this intuition, it is not the falsity of the proposition that should explain the wrongness of Ferdinand’s assertion; rather, the lack of appropriate evidence. We now have a second candidate formulation of condition C: do not assert a proposition *unless you have appropriate evidence*. Yet, some other people might intuit that Ferdinand’s assertion is at fault in both ways: you should neither assert what is false, nor what is not supported by appropriate evidence. If this is the case, then perhaps condition C is that a speaker should *only assert what he knows* (as knowledge requires both truth and appropriate evidence).

These are but a few possible solutions to the question raised by the hypothesis. There is a vast literature on the issue, ripe with different accounts of what makes an assertion permissible. For the moment, I will not enter the intricacies of this debate, nor will I attempt to explain which account offers a better explanation of the norm regulating assertion. All I aimed to clarify is that the hypothesis that assertion is subject to a single epistemic norm has the potential to explain what is wrong with assertions that are not lies, but that are nonetheless false or unwarranted. This will be the explanandum of the second part of the dissertation. In those pages, I will assess the solidity of the hypothesis to which the scholars involved in the debate subscribe, and attempt to establish which kind of condition C should be.

### 3. Outline of the Dissertation

As already mentioned, this dissertation is divided into two parts. Part A discusses problems concerning lying, and part B problems relating to the norm of assertion.

**Part A** opens by introducing the philosophical debate about the definition of lying. In the contemporary discussion, two factions are opposed: *deceptionist*, who believe that lies necessarily aim at deceiving someone, and *non-deceptionist*, who deny this necessary condition. **Chapter 2** presents some known arguments against deceptionist definitions, and supplements them with novel arguments. It concludes by responding to some recent objections presented by deceptionists in reply to these arguments. **Chapter 3** turns to criticisms of the other faction, non-deceptionism. It opens by recapitulating some known counterexamples that affect each of the most popular non-deceptionist definitions. It then proceeds to show that all these accounts have a common defect in common: they cannot deal with three ‘speech-act theoretic’ puzzles about lying – puzzles that involve lies performed by means of a speech act other than assertion. To find a solution to these puzzles, I turn to speech act theory, with the aid of which I develop a non-deceptionist definition that avoids the objections to which all the other definitions are subject. In order to further refine this account, **Chapter 4** tackles the problem of defining insincerity. It is divided into four main sections. The first extends the insincerity conditions to speech acts other than assertion, showing how insincerity is not necessarily a matter of beliefs. The second tests this account against ordinary speaker intuitions, finding support for the proposed account. The third deals with the problem of graded insincerity. Speakers can be more or less confident in what they say, and communicate a higher or lower degree of confidence in what they say. Insincerity involves some discrepancy between the belief held by the speaker and the belief communicated by its utterance: my account proposes a system to model this issue and define which degree of discrepancy is required to call something a lie. In the closing section, I bring together all the findings, to develop a unified definition of lying.

**Part B** begins by introducing the norm of assertion hypothesis. According to this hypothesis, there is one norm that defines whether a proposition can be asserted or not, and this norm takes the form: “assert that *p* only if *p* has C”. In order to identify which propositions are assertable, philosophers have to determine exactly which property is C. **Chapter 5** puts some pressure on this hypothesis, and more specifically on the idea that the envisaged norm is *constitutive* of assertion. It shows that different authors writing on the norm of assertion have interpreted this claim in different ways. I argue that no matter which interpretation is chosen, the idea that the envisaged norm is constitutive of assertion is misguided. Once one recognises this, the assumption that assertion is subject to *only one* norm of this kind loses its appeal, and so does the idea that assertion can be defined as the only speech act that is only subject to this rule. **Chapter 6** reviews two ways of fleshing out the norm of assertion hypothesis: *factive* and *non-factive* accounts. The former maintain that only true propositions are assertable, whereas the latter deny this, and argue that some false propositions are assertable. Factivist positions are subject to known counterexamples (*unlucky assertions*), but they are generally defended with a reasonable argument: that only a factive account can explain why false assertions are incorrect and liable to criticism. I deny this claim, showing that the same data can be explained by taking truth to be the *purported aim*, rather than the rule, of assertion. I conclude by arguing that, in the light of this data, the account offering the best prediction will be one featuring one or more non-factivist rules, paired with the view that assertion aims at truth. In **Chapter 7** I summarise the key ideas that I defended throughout the thesis.

**Part A- Lying**

# A. The definition of lying

Each of the two parts of this dissertation will feature an introductory chapter like the present one – labelled with a letter, instead of a number, to indicate its different nature. In this introductory chapter, the philosophical debate on the definition lying is presented. After a short historical introduction, I introduce the *classic definition* of lying. This definition involves two key conditions: the *statement condition* and the *insincerity condition*. Each is briefly introduced and explained with the help of examples. I then proceed to consider the contemporary debate, that involves an opposition between two opposite factions: *deceptionist* and *non-deceptionist* (or *assertion-based*) definitions. After presenting each of them, I set the ground for the discussion in the further chapters.

### 1. The Classic Definition of Lying

As Augustine gracefully points out in his *De Mendacio* (possibly the first work to systematically discuss this issue), the question of lying “is, indeed, very full of dark corners, and has many cavern-like windings, whereby it oft eludes the eagerness of the seeker; so that at one moment what was found seems to slip out of one's hands, and anon comes to light again, and then is once more lost to sight”. Augustine is stressing an important truth: defining lying (and the morality thereof) is a more difficult task than it seems at first sight. As a matter of fact, the history of philosophy has proved him right: in almost two millennia of discussion, scholars have not reached a consensus about which definition best captures the concept of lying.

Nevertheless, even though there is no agreed upon definition of lying, there are some firm points of agreement. If our discussion of the definition of lying has to start somewhere, it probably has to start from such firm points. Three in particular seem to be resistant to even the strongest scepticism. First, lying is an intentional act: there is no such thing as lying unintentionally. Second, lying requires making a statement: in order to tell a lie, you have to produce some linguistic token. Third, lying requires making an insincere statement: there is no such a thing as a sincere lie, because lying involves stating something that you do not believe.

From these three firm points, we can get to a first tentative definition of lying:

**Classic definition:**

(CD) To lie is to intentionally state something that you believe to be false

This definition gets most simple cases right. Consider an obvious example of lying: if Pinocchio tells Geppetto “I went to school this morning” even if he did not go, intuitively he has lied. Since Pinocchio is intentionally telling Geppetto what he believes to be false, the definition correctly dubs this case a lie. Now, suppose that later that day, Pinocchio is asked where Geppetto is. Pinocchio has seen Geppetto at the workshop, so he replies “My dad is at the workshop”. It turns out that Pinocchio is wrong, as Geppetto snuck out of the workshop to visit his lover in secret. In this case, Pinocchio has said something false, but his statement is sincere: intuitively, he has not lied. Once again, the intuition is tracked by the definition, that rules out this case as the statement is not believed to be false. At first blush, the classic definition deals correctly with simple cases of lying, and correctly distinguishes them from other false or deceptive utterances.

I called this definition ‘classic’. This term may seem unorthodox to some readers, as philosophers (*e.g.* Lackey 2013, Mahon 2015) often have in mind a different definition when they talk about the *standard* or *traditional* definition of lying (which I will soon introduce as the ‘deceptionist definition’). I purposefully chose the term ‘classic’ instead of ‘standard’ or ‘traditional’ to acknowledge this departure from orthodox terminology. Such departure is justified by two reasons*.* The first is that this was very likely the first philosophical definition to appear in the literature, and remained the standard definition for at least a thousand years: introduced by Augustine (DM, IV AD), it is still considered the standard view in the works of Aquinas (ST, XIII AD) and Peter Lombardus (SEN, XIII AD)[[10]](#footnote-10). The second reason is that, while there is no consensus as how to define lying, the overwhelming majority of scholars agree that intentionally stating what you believe to be false is a *necessary* condition for lying: you cannot lie unless you meet the conditions stated by (CD). In other words, the current consensus is that (CD) offers an accurate *characterisation* of lying, even if it is arguably not an accurate enough *definition* – because meeting the conditions stated in (CD) may not be sufficient for a statement to qualify as a lie.

The current debate on the definition of lying mainly revolves around the question of which additional condition is best suited to make the definition meet sufficiency. Before entering the current debate, however, it is worth familiarising ourselves a bit more with the points of agreement set by the classic definition. More specifically, in the next two sections I will discuss in detail the two main requirements posited by the classic definition: the statement condition and the (intentional) insincerity condition.

#### 1.1 The Statement Condition

*You! You chameleon! Bottomless bag of tricks! Here in your own country would you not give your stratagems a rest or stop your spellbinding for an instant?*

Homer, *Odyssey,* XXIV

In his *Parerga e Paralipomena* (1851/1974:538), Arthur Schopenhauer writes that “there is in the world only one mendacious and hypocritical being, namely man. Every other is true and sincere, in that it frankly and openly declares itself to be what it is and expresses itself as it feels”. Unlike animals, whose ingenuity and transparency is fascinating to us, the degenerate human tendency to lie “stands as a blot on Nature”. Schopenhauer’s severe verdict is certainly inaccurate, given that plants and animals are capable of incredibly complex forms of deception. However, in his observation we can find a grain of truth: arguably, non-humans cannot lie, at least if lying requires *telling* a lie – that is, uttering a linguistic token that we believe to be false.

The intuition that lying requires linguistic abilities is reflected by the classic definition of lying, according to which lying involves stating a proposition, or saying something. Following Mahon (2015), I call this the *statement condition* for lying. The statement condition captures the intuitive difference between lying and other forms of deception. Pace to Schopenhauer, countless examples of deception can be found in nature to illustrate this distinction. For instance, the orchid *Cryptostylis erecta* is pollinated by the so-called orchid dupe wasp (*Lissopimpla excelsa*), the males of which mistakethe flower parts for female wasps, and copulate with them. While there is a sense in which the wasp is *deceived* by the orchid’s shape, it would be erroneous to say that the orchid *lied* to the wasp. The statement condition thus captures the intuitive distinction between lying and simple deception.

A few authors reject the difference between deception and lying [[11]](#footnote-11). However, this leads them to conclusions that are rather counterintuitive. As *a reduction ad absurdum*, it will suffice to remark that Smith (2004), who rejects the statement condition, lists as lies “breast implants, hairpieces, fake orgasms and phony smiles, as well as age-concealing make up and deodorants that disguise our scent”. While these cases usually involve attempted deception, they are clearly not lies. Claiming that wearing deodorant is lying can strike as bizarre, and illustrates the counterintuitive consequences of rejecting the statement condition.

The view that lying involves stating something can be traced back to Augustine’s seminal work. In his Contra Mendacio (XII) he writes that “a lie is a false signification by words”. Aquinas (ST, q110), who knew Augustine’s work thoroughly, rightly specifies that this does not mean that lying is necessarily a matter of verbal communication:

As Augustine says [DDC, II], words hold the chief place among other signs. And so when it is said that "a lie is a false signification by words," the term "words" denotes every kind of sign. Wherefore if a person intended to signify something false by means of signs, he would not be excused from lying

These observations are the right track: there are cases of lying that do not involve the use of words. One can lie by using any sort of conventional signals, or combinations thereof. For instance, you can lie by using body gestures that have conventional meaning (as in nodding with your head to agree), or using smoke signals, and so on. To capture all these cases, scholars working on lying usually refer to Chisholm and Feehan’s (1977:150) definition of statement:

**Definition of Statement** (Chisholm & Feehan)

S states that p to A iff

(a) S believes that there is an expression E and a language L such that one of the standard uses of E in L is that of expressing the proposition p;

(b) S utters E with the intention of causing A to believe that he, S, intended to utter E in that standard use

In other words, a lie has to be a linguistic token that is believed to expresses a proposition, and that is believed to do so in virtue of some linguistic convention. This formulation of the statement condition is preferable to Augustine’s, as it does not seem that “every sign” used to signify something false can be used for lying. Consider the example of wearing a lab coat to pretend that you are a scientist, or a ring to pretend that you are married. In these cases, you use a sign to communicate something false, but you are not lying. The statement condition proposed by Chisholm and Feehan correctly rules out these cases. To go back to Schopenhauer’s erroneous claims about nature’s intrinsic sincerity, another example of deceptive signals that are not lies is found in animal signalling. Many animals produce signals that are associated with a stimulus. Sometimes, however, they use such signals in the absence of the stimulus, for deceptive purposes. For instance, *Lanio Versicolor* sentinel birds are known to produce alarm calls in the absence of predatory birds, in order to scare their conspecifics when the competition for food is high (Munn 1986). According to the statement condition, and consistently with intuitions, these alarm calls are deceptive, but are not lies[[12]](#footnote-12).

Lastly, some authors (Siegler 1966, MacCormick 1983, Fallis 2010; 2013; 2014, Meibauer 2011; 2014, Saul 2012, Stokke 2013; 2017, Viebahn 2017) prefer to use the term ‘saying’ in place of ‘stating’. In the literature, the two terms are understood to be synonyms: they both indicate the utterance of a meaningful declarative sentence. However, it should be noted that a minority of authors that use the term ‘saying’ (Saul 2012, Viebahn 2017) do this to mark their commitment to a slightly different version of the statement condition – more specifically, a Gricean account of what it means to *say* something. For the purpose of this dissertation, however, we can safely ignore this subtle distinction. Consequently, in this dissertation I will treat ‘saying’ and ‘stating’ as synonymous.

#### 1.2 The Insincerity Condition

*Non enim omnis qui falsum dicit mentitur*

*si credit aut opinatur verum esse quod dicit*

Augustine, *De Mendacio*, 3.3

According to the insincerity condition, the speaker has to *believe* that what is said is false. I call this condition the *insincerity* condition, because it captures the difference between an insincere utterance (that you believe to be false) and a mere mistake (that you believe to be true, but turns out to be false). Going back to the previous example, if Pinocchio mistakenly believes that Geppetto is in the workshop and voices his mistake, he says something false, but he does not lie: while incorrect, his utterance is sincere. The insincerity condition captures this intuition: saying something false does not amount to lying, unless the falsehood is also believed to be false.

Importantly, requiring that the statement is believed to be false is not yet requiring that the lie must be false. Most scholars deny that actual falsity is necessary for lying[[13]](#footnote-13). However, a few authors (Grotius RWP:1209, Benton forth, Turri & Turri 2016) have suggested that falsity of the statement, in addition to *belief* in its falsity, is required for lying. For these authors, if your believed-false utterance turns out to be true, you have not lied. People seem to have different intuitions in this respect. For our purposes, it will suffice to say that the classic definition gives the opposite verdict, and counts inadvertent truth-telling as lying.

The insincerity condition also grounds the intuitive distinction between lying and merely misleading. A classic example of misleading that falls short of lying is found in the hagiography of St. Athanasius (MacIntyre 1994: 336):

Persecutors, dispatched by the emperor Julian, were pursuing [Athanasius] up the Nile. They came on him travelling downstream, failed to recognise him, and enquired of him: “Is Athanasius close at hand?” He replied: “He is not far from here.” The persecutors hurried on and Athanasius thus successfully evaded them without telling a lie.

In this example, Athanasius is attempting to deceive is pursuers by implying that he is not Athanasius. While deceptive, his carefully phrased sentence is true; intuitively, it does not constitute a lie. The insincerity condition makes sense of this intuition: it predicts that this is not a lie, given that Athanasius does not believe that what he has said is false.

Finally, according to the classic definition the speaker has to *intentionally* say what he believes to be false. This requirement of intentionality is found in some classic texts; here is how Griffiths (2004:27) introduces it:

Them is an internal fact and an external fact. The internal is what's in the mind (animus—Augustine often also uses "heart," cor, or "mind," mens, for the same purposes) and the external is what's said or communicated in some other way—by gesture or expression or some other nonverbal sign. Lying happens when the two are intentionally separated. (my emphasis)

One way of understanding the requirement that the speaker *intentionally* says what he believes to be false is for it to be a refinement, or specification, of the insincerity condition. To be insincere, in this sense, is to *intentionally* establish a discrepancy between what you believe and what you say. To some, the adverb ‘intentionally’ may seem redundant here. After all, unintentional falsities (*i.e.* mistaken but sincere statements, as in Pinocchio’s statement that Geppetto is in the workshop) are already ruled out by the requirement that the speaker says something he *believes* to be false. However, the intentionality constraint plays an important role here: it rules out other species of accidental falsehoods from the definition of lying and insincerity.

Linguistic mistakes (cases in which when the speaker misspeaks, or gets confused about the meaning of words) offer a first example of such accidental falsehoods. Saul (2012:14, see also 2012:16) has a rather amusing example:

Anna, an English rock climber, wanted to tell her [Mexican] colleagues that many people in England climb without ropes. So she uttered (2):

(2) En Inglaterra hay mucha gente que escala sin ropa

(2) actually means that in England there are many people who climb without clothes [ropa]. This claim is false, but Anna did not lie; she accidentally said something false, through a linguistic error.

A definition that only requires a liar to state a believed-false statement would count (2) as a lie: Anna has said that people in England climb without clothes, and she believes this to be false. However, Anna has not said what she believed to be false *intentionally*, so that this case is not counted as a lie by the classic definition. The intentionality requirement is thus indispensable to maintain the distinction between lies and malapropisms.

Self-deception also generates examples of accidental falsehoods that can only be addressed by the intentionality constraint on insincerity. Self-deception is a psychological condition whereby one convinces oneself of the truth of something that one knows to be false (hence the deception), and has no awareness of being so deluded. In other words, when you are self-deceived about *p*, you believe that you believe that *p*, but you do not in fact believe that *p*. Ridge (2006:488–9) offers an example:

Bob believes that he believes his mother loves him but actually does not believe that she loves him. In fact, Bob believes his mother hates him. [...] Suppose we ask Bob whether his mother loves him and he says, ‘‘Yes, of course she does’’.

In this case, Bob is saying something that he believes to be false. Nonetheless, he is not saying something he believes to be false *intentionally*. Once more, the intentionality constrainton insincerity is doing the heavy lifting in distinguishing lying from other accidental falsehoods: were we not to require that the “separation between mind and words” was intentional, we would incorrectly count these cases as lies.

### 2. The Current Debate: Deceptionist vs Non-Deceptionist accounts

There should be no doubt at this point as to the exact meaning of (CD) within the debate on lying, and the arguments that support it should be clear.

**Classic definition:**

(CD) To lie is to intentionally state something that you believe to be false

Now, I have already mentioned that while there is consensus that (CD) offers an accurate *characterisation* of lying (it is correct about what is required for lying), it is not taken to be an accurate *definition*, because meeting the conditions stated in (CD) is not sufficient for a statement to qualify as a lie.

This criticism of the classic definition is grounded: there are indeed statements that are believed to be false and are not lies. Believed-false statements that are not lies include *ironic statements; fictional statements* (e.g. uttered on stage or written in a fictional novel); *jokes; teasing remarks; hyperboles; metaphors; euphemisms*; and so forth. These kinds of statements, that I will call *non-assertoric falsehoods,* represent solid counterexamples to the classic definition: they meet (CD), but intuitively are not lies.

To see this, let us consider an example of non-assertoric falsehood, a fictional statement. Imagine that an actor on stage utters:

1. I am Ubu, Prince of Podolie, Duke of Courlande, Earldom of Sandomir, Margrave of Thorn

The actor is saying something that he believes to be false, but he is clearly not lying – he is just pretending to be Ubu for the sake of the play. Since the classic definition has no resources to rule out these cases, it must be incorrect. To be sure, this is quite an important failure: it means that the definition gives incorrect verdicts in a very wide variety of cases, involving virtually every figure of speech that can be literally false.

Upon the failure of the classic definition, there are essentially two strains of definitions that are able to solve this problem and distinguish between lies and other *non-assertoric falsehoods* as non-lies: *deceptionist definitions* and *non-deceptionist, assertion-based definitions.* Deceptionist definitions expand (CD) by introducing the further condition that the speaker must intend to deceive his audience. This amendment deals with *non-assertoric falsehoods* by excluding them in virtue of the fact that they are not intended to deceive. For instance, the actor’s utterance that (1) is not counted as a lie, because the actor is not attempting to deceive his audience. Broadly, deceptionist definitions[[14]](#footnote-14) are phrased as follows:

**Deceptionist definitions:**

S lies to A iff:

(a) S states that p

(b) S believes ¬p

(c) S intends A to believe p

In recent years, there has been growing consensus that these definitions are incorrect. Their key problem is that the ‘intention to deceive condition’ (c) exposes the definition to several counterexamples (that will be discussed extensively in the next chapter).

Are there alternative ways to amend the classic definition? The most influential alternative is to require that the speaker *genuinely asserts* that *p*. This amendment deals with *non-assertoric falsehoods* by excluding them in virtue of the fact that they are not genuinely asserted. For instance, the actor’s utterance on stage does not count as a lie because the actor is not genuinely claiming that he is King Ubu, but merely pretending to claim for the sake of the play. Definitions that follow this strategy are dubbed *non-deceptionist*, because they reject the intention to deceive conditions, and *assertion-based –* because, unlike the classical definition (which is also *non-deceptionist*), they introduce the further requirement that the relevant proposition is asserted. More formally, assertion-based definitions read:

**Assertion-based (non-deceptionist) definitions:**

S lies to A iff:

(a) S says p

(b) S believes ¬p

(c’) S asserts that p

In sum, the contemporary debate on the definition of lying articulates around

which putative additional condition is required to amend the classic definition of lying: some authors believe that the speaker has to *intend to deceive* the audience, other that he has to genuinely *assert* the relevant proposition. Chapter II will deal with the deceptionist accounts, and Chapter III will discuss their most prominent alternative, assertion-based accounts.

# II. Deceptionist definitions of lying: a critique

Yet we must observe, there are many people to be met with, who will utter lies, not with any malicious intention, or for any advantage to themselves, or to injure others either in their property or in their reputation, but merely for the pleasure of the lies themselves: as you meet with some people, who will be tippling every moment, not to quench their thirst, but merely from a sottish habit of drinking.

Giovanni Dalla Casa, Galateo: Or, A Treatise on Politeness and Delicacy of Manners.

Deceptionist accounts of lying require that, in uttering an insincere statement, the speaker must intend that statement to deceive the hearer. This chapter introduces and criticises the intention to deceive condition, in order to prove that it is neither necessary nor sufficient (jointly with the other conditions) for lying.

## 1. The intention to deceive condition: an introduction

According to deceptionist definitions, lying is more than an insincere statement: it is an insincere statement that is intended to deceive the addressee. Call this requirement the ‘intention to deceive condition’ (henceforth, IDC). Authors who defend deceptionist accounts of lying are motivated by two parallel and mutually supportive intuitions. On the one hand, there is the simple idea that having a deceptive intention is part of the essence of lying: it is a conceptual truth that to lie is to aim to deceive someone. On the other hand, there is the idea that there are believed-false statements that are not lies, and all these statements have one feature in common, namely that they are not intended to deceive.

I have already introduced this category of believed-false statements that are not lies, and that pose a threat for the classic definition. I call them ‘non-assertoric falsehoods’: fictional statements; jokes; teasing remarks; hyperboles; metaphors; euphemisms; and so forth. These are statements that are believed to be false, but not intended to deceive and therefore not lies. One important selling point of the IDC is its success in singling out non-assertoric falsehoods: this family of believed-false statements is not intended to deceive, so that the IDC correctly excludes them from the definition.

It is worth noting that IDC does not require that the audience be deceived. This correctly tracks the intuition that, while ‘deceive’ is a success verb (Ryle 1949: 130), ‘lie’ is not (Mannison 1969: 135; Wood 1973: 199; MacCormick 1983:9, fn23). The IDC only requires that the speaker intend to deceive the hearer.

Most philosophers interpret the IDC as indirectly requiring that the speaker thinks that it is possible for the statement to deceive audience. Arguably, having an intention to φ requires thinking that φ-ing is possible (Grice 1989:98, Fallis 2015). If this is correct, the IDC correctly excludes the possibility for a rational speaker to lie to someone who cannot have mental states and specifically beliefs – like an object, a computer, or a person in a persistent vegetative state. Similarly, this excludes the possibility to lie to someone you know to be unable to understand your statement, like an animal, a toddler, a foreigner, a deaf person, etc.

In recent years, a number of philosophers have challenged the IDC, generating a rich debate on whether intending to deceive is necessary for lying. This chapter offers a critical review of the relevant literature. It is organised as follows. Each section discusses one of the three kinds of strategies that can be employed to challenge the IDC. Section 2 shows, following Fallis (2010), that each proposed phrasing of the IDC (including 1) falls victim to compelling counterexamples, and that no existing phrasing is immune to counterexamples. Section 3 introduces a new difficulty for the IDC: the fact that it yields wrong verdicts on statements with no assertoric force that are intended to deceive. Section 4 presents some classic and more direct counterexamples: lies in which the speaker does not intend to deceive the hearer. Section 5 presents and deflects some recent reply to these counterexamples.

## 2. Different versions of the intention to deceive condition

### 2.1. The ‘broad’ IDC

There is disagreement between proponents of deceptionist accounts as to which phrasing of the IDC better tracks our intuitions on lying. I will start by considering the broadest phrasing (Mannison 1969:133, Newey 1997:100-02):

**Broad Intention to Deceive Condition**

IDC1 S intends his statement to deceive A

This condition counts any attempt to deceive the hearer as relevant – as long as it is meant to be caused by the statement. This means that IDC1 correctly captures insincere statements that attempt to make the hearer believe the very propositional content of the statement. Moreover, IDC1 also captures lies that attempt to make the hearer believe something related to, but different from, the propositional content of the statement. For example, IDC1 correctly rules in statements that attempt to trick the hearer into thinking that the speaker incorrectly believes the propositional content of the statement to be true (cf. Doxastic Misdirection, section 2.2).

Some philosophers, however, have argued that IDC1 is too broad, as it incorrectly captures insincere utterances that attempt to deceive the hearer about a proposition that bears little or no relation to the statement. Take the following example from Fallis (2014):

Theatrical Pose

“Suppose that I decide to try to convince a new acquaintance at a party that I am a professional actor (rather than an academic). So, I take a theatrical pose and intone, “This is I, Hamlet the Dane”. In making this statement that I believe to be false, I clearly intend to [deceive my acquaintance] (viz., about my being an actor). But I am not lying because I do not intend to be deceptive about what I actually say (viz., about being named Hamlet or about being Danish).”

It is pretty straightforward that, in the example, I am not lying. The problem with this formulation of the IDC is that it includes intents to deceive that are totally disconnected from the content of the statement. The example shows that IDC1 is too broad, as it counterintuitively captures statements whose intent to deceive has little or nothing to do with what is said. Hence, IDC1 needs to be abandoned.

### 2.2. The ‘rigid’ IDC

The simplest way to avoid the problems of condition IDC1 is to lock the intent to deceive to the very propositional content of the utterance (Kupfer 1982:104, Williams 1996:147, Williams 2002:96, Mahon 2008:228). This is the narrowest version of the IDC, and is probably the most common in the literature. It can be spelled as follows:

**Rigid Intention to Deceive Condition**

IDC2 S intends his statement that p to deceive A into believing that p

This ‘rigid’ version of the IDC avoids the most important problem of IDC1: it correctly rules out cases like Theatrical Pose, because they involve deception about something other than the propositional content. More generally, it is able to track the intuition that there must be some link between the deceptive intent and the content of the statement. Nevertheless, IDC2 is vulnerable to several counterexamples.

A first problem with IDC2 is that, being so restrictive, it captures only statements meant to deceive about their very propositional content. As a result, in some cases IDC1 fares better than IDC2. A classic example is doxastic misdirection (Fallis 2010). If you say something you believe to be false to convince your hearer that you incorrectly believe it to be true, IDC2 is not satisfied (because you want the hearer to think that you are wrong) but it seems that you are still lying (after all, you want to convince him that you believe *p,* which is false). To make this point clearer, consider the following example (Fallis 2010: 3):

Doxastic Misdirection

“A child fears that if his parents were to discover that he no longer believes in Santa Claus, they would stop giving him Christmas gifts. So, opening his gift he utters: “Look! Santa brought me a new bicycle,” to make them think that he still believes in Santa Claus. The child is aware that his parents know that the statement is false, and he clearly does not intend them to believe that the statement is true – i.e., that Santa Claus brought him the bicycle. [...] Thus, the speaker is not lying according to [deceptionist definitions]. Nevertheless, the speaker seems to be lying”.

Several authors (Chisholm & Feehan 1977: 151-52, Newey 1997:100, Fallis 2010, Mahon 2015:1.4) argue that in these cases the speaker is clearly lying, and Doxastic Misdirection seems indeed a straightforward case of lying. Since IDC2 incorrectly excludes these cases, it needs to be revised.

### 2.3. The ‘believed sincerity condition’ and the disjunctive account

An available move is to loosen IDC2 so that it includes these cases. We could simply require that the speaker intend his audience to believe that he is speaking sincerely, *i.e*. that he intends the audience to believe that he believes what he says. Let us call this condition the *believed-sincerity condition* (BSC):

**Believed-Sincerity Condition**

(BSC) S intends his statement to cause A to believe that S believes that P.

Often, authors embrace both IDC2 and BSC (Simpson 1992: 625; Frankfurt 1999: 5, Faulkner 2007: 527-8, 2013), rather than BSC alone. Roughly, the idea is that you cannot expect the hearer to believe what you say if he believes that you disbelieve it – hence, intending to deceive (IDC) requires that you also attempt to make your hearer believe that you are sincere (BSC).

However, embracing both IDC2 and the BSC is not desirable, since the revised definition would still exclude cases of doxastic misdirection. A more plausible alternative is to require only BSC to be satisfied. If the previous considerations are correct, and intending to deceive requires intending that the audience believe that the liar is sincere, when IDC2 is satisfied also BSC is satisfied. BSC alone would then capture all the lies captured by IDC2, plus cases doxastic misdirection, thus meeting our desiderata.

At this point, counterexamples get a bit more complicated. Against BSC, it is possible for the liar to have the intent to make the hearer believe that *p* without having the intent to make the hearer believe that the speaker believes that *p.* In other words, there are cases of lying in which IDC2 is satisfied and the BSC is not. Consider the following case, inspired by Fallis (2010: 5)

Two Colours Party (Double Bluff)

Mario has been invited to a party where one is allowed to dress either in red or in green. He bought a shiny red suit, but wants to trick his friend Luisa into thinking that he bought a green one instead. He is convinced that Luisa thinks (a) that Mario wants to respect the rule about the colours (i.e. that he either bought a green suit or a red one); (b) that Mario is affected by daltonism, and that people affected by daltonism always gets green and red mixed up – namely, that Mario will believe he has a green suit if he has a red one, and vice versa; (c) that Mario will lie about the colour of his suit. Being convinced that Luisa thinks (a-c), Mario tells her “I just bought a green suit!”. He hopes that, since Luisa believes that he is both wrong and insincere about the colour, she will conclude he has bought a green suit.

In the example, Mario tells Luisa something he believes to be false, intends her to believe what he says, but does not intend Luisa to believe that he believes what he says. Therefore, IDC2 is satisfied and BSC is not. This case, though complex, is intuitively a case of lying, one that BSC fails to capture.

Mahon (2008: 220) has suggested a revised condition that deals correctly with both Doxastic Misdirection and Double bluff. The proposed condition is a disjunctive condition: it requires that either IDC2 or the BSC be satisfied. More formally:

**Disjunctive Intention to Deceive condition**

(IDC3) S wants A to believe that p is true or S wants A to believe that S believes p or both

The disjunctive condition avoids the previous counterexamples. A first difficulty with IDC3 is that it presents a disjunction, and introducing a disjunction in a definition carries the risk of bringing together under the same label what in fact are two phenomena, and is often a symptom of ad-hoc revisions rather than theoretically justified moves (Fallis 2010, Kingsbury & McKeown-Green 2009, 578-81).

More importantly, IDC3 fails to deal with other cases of bluffing. The philosophical literature is rich with examples of bluffing where the speaker is lying, there is intent to deceive, but IDC3 is not satisfied (Newey 1997; Moran 2005:17; Mahon 2008; Fallis 2010). A classic example is found in S. Augustine [DM] (quotation from Simpson 1992: 628):

Bandits (Simple bluff)

“A man [call him Simplicius] knows that there are bandits on a certain road and fears that his friend may take that road. The man knows that his friend does not trust him, and so says that there are not bandits on the road, hoping that his friend will come to believe that there are”.

Here Simplicius states what he believes to be false, with the intent to make his friend believe that (i) what he says is false and that (ii) he believes that what he says his false. So neither the BSC nor IDC2 are satisfied. However, even if Simplicius has the benevolent intent of saving his friend, it seems that he is lying, because he is speaking insincerely and with a deceptive intent (he is attempting to make his friend believe that he wants him to believe that the statement is true, while in fact he wants him to believe that it is false).

There seems to be consensus that, even though Simiplicius’ statement in the Bandits scenario is told with good intentions, it is a lie. But it is certainly an option for the deceptionists to deny this intuition, and insist that bluffs like this are not genuine instances of lying, because no intention to deceive is involved. Perhaps this reply is question begging; but even to these demanding philosophers we can offer a satisfying counterexample. There are counterexamples to IDC3 that *do involve* an intention to deceive – just not the intention to deceive captured by any of the previous definitions. For instance, consider the following variations on the Two Colours Party scenario (inspired by Fallis 2010):

Three Colours Party (Double Bluff with a third option)

The scenario is the same as Two Colours Party, but with two differences: (1) Mario decided not to respect the rule about the colours, and chose a third colour for his suit – white; (2) Luisa does not believe Mario is affected by daltonism – she only thinks that Mario wants to deceive her. Mario wants Luisa to believe he bought a red suit rather than a white one, so he tells Luisa “I just bought a green suit!”. He hopes that, since Luisa believes he has either bought a red or a green suit, and that he is insincere, she will conclude that he has bought a red suit.

In the example, Mario states a believed-false proposition (*the suit is green*) with the intention to make Luisa believe another believed-false proposition (*the suit is red*). Clearly, there is intent to deceive. But IDC3 is not satisfied, because Mario neither intends Luisa to believe he is speaking sincerely, nor does he intend her to believe the propositional content of his statement to be true. Since Bandits and Three Colours Party are cases of lying and both IDC2, the BSC and the IDC3 fail to track this intuition, we need to abandon these formulations of the IDC.

None of the phrasings of the IDC considered in this section successfully tracks our intuitions on lying. Even though several philosophers share the intuition that lying requires intending to deceive, it seems that none of the definition on the market is able to deal correctly with each and every case. In the next two sections, we will consider two further arguments: that deceptionist accounts incorrectly include non-assertoric utterances, and that deceptionist accounts incorrectly exclude lies that are not intended to deceive.

## 3. Against sufficiency: non-assertoric falsehoods that intend to deceive

Poets themselves, tho' liars by profession, always endeavour

to give an air of truth to their fictions

David Hume, *A Treatise on Human Nature*, I, III,10

Now for the poet, he nothing affirms, and therefore never lieth. For, as I take it, to lie is to affirm that to be true which is false. […] But the poet (as I said before) never affirmeth […]. And therefore, though he recount things not true, yet

because he telleth them not for true, he lieth not

Sir Phillip Sydney, An Apology for Poetry

Faced with the counterexamples presented in section 2.2, the defender of the IDC could once again bite the bullet and argue that in *all* the bluff-based examples (Bandits, Two Colours Party, Three Colours party) the speaker is not lying. Even if most authors share the opposite intuition (Plunze 2001:185-7; Mahon 2008; Fallis 2010), some (like Faulkner 2007:536-7; 2013) follow this strategy and contend that bluff cases are correctly excluded by deceptionist accounts.

Even if one accepts this view, there is a further problem with deceptionist accounts. In section 1 I claimed that a selling point of deceptionist accounts is that they successfully single out all the believed-false statements that are not lies, as these statements (that I dubbed *non-assertoric falsehoods*) all lack intent to deceive. However, it seems that at least some believed-false statements can be intended to deceive without thereby counting as lies. If this is true, deceptionist accounts are too broad. To see this, consider the following example of a believed-false statement uttered with deceptive intent (for a somewhat similar example, cf. Fallis 2009:55-6):

Historical fiction

George is a Scottish writer, well known for his faithful representation of Scottish history. His works are historical novels of the realist genre. Apart from the lives of his characters, the historical references in his novels are predominantly accurate, taken to be so by the readers, and intended to be interpreted as such. George is now writing a new novel, and this time he wants to convey a vision of history that will favour the independentist cause he supports. He thus cunningly sprinkles his novel with false historical claims meant to strengthen Scottish identity, like: “Scots begun to wear the kilt at the beginning of the XIth century”. He intends these statements to be interpreted as statements about the real world, believes them to be false, and intends his audience to believe them to be sincere and true – which is likely to happen, given that his previous novels were predominantly accurate and sincere. However, given that the novel is partly fictional, George is not warranting the truth of his statement: he is merely suggesting that the statement is true. For this reason, it seems that it would be odd to regard his statement as a lie.

In Historical fiction, the speaker (to be exact, the writer) states what he believes to be false and attempts to deceive the reader. The example satisfies the intention to deceive condition, no matter how it is phrased: all deceptionist accounts deem George’s utterance a lie (or, at the very least, we can conceive variants of the example satisfying every version of the IDC). However, it seems that George is not lying, since his statement occurs within a fictional work, and George is only suggesting that what he wrote is true – rather than explicitly taking responsibility for its truth (for a similar example, cf. Reptilian President in III.1.2).

Sir Phillip Sidney’s famous quote (“the poet, he nothing affirms, and therefore never lieth”) mentioned in the opening of this section tries to get the same point across: fictional statements are not asserted and therefore not lies. Philosophers tend to share this view[[15]](#footnote-15): we must distinguish between ‘serious’, assertive discourse (for the truth of which a speaker takes full responsibility) and fictional, ‘etiolated’ (the term is from Austin 1962/1975) discourse (for the truth of which the speaker cannot be held responsible – even when, like in Historical fiction, there is a regular connection between the author’s and actual facts). And philosophers tend to agree that a fictional utterance (i.e. an utterance occurring within the context of a fictional work) can suggest that a proposition is true, but it cannot genuinely assert it (Mikkonen 2010). Now, in Historical fiction George merely suggests, or implies, that the historical facts related in his novel are true rather than imaginary. Given that the truth of their propositional content is only implied, the speaker is not liable to be blamed or resented in the same sense that he can be blamed for having lied. Therefore, a satisfactory account of lying should classify these utterances as misleading statements rather than lies[[16]](#footnote-16).

The problem identified with fictional statements is only the tip of the iceberg, as it is shared by any utterance that is believed to be false, intended to deceive the audience but not genuinely asserted. Ironic statements, euphemisms and jokes are also incorrectly classified as lies by deceptionist accounts, whenever a relevant intention to deceive obtains. For instance, consider the following example:

Ironic Double Bluff

At a meeting between Mafia gangsters, Neil tells to Jack: “Yesterday I finally managed to lure Ludmilla into my bed!”. Jack knows that Neil is obviously telling tales, so he replies: “Ahaha! Of course, you laid with Ludmilla! And I am a cop!”. Jack is not a cop, but he knows that Neil suspects that he is an undercover cop. He hates Neil, and wants to corroborate Neil’s suspicion in order to make him look paranoid in the eyes of their accomplices. He expects Neil to follow this line of reasoning: “Jack chose this reply because he is indeed a cop, and he is only trying to make the thought of him being a cop look ridiculous”. In this case, the statement aims to deceive Jack into believing what is said (that he is a cop) and is believed to be false, so that it is a lie according to deceptionist accounts. However, the speaker is speaking ironically, so that it would be odd to dub this a lie.

The proposed example is clearly not a case of lying. However, Neil is saying what he believes to be false with a deceptive intent that satisfies IDC1, IDC2 and BSC[[17]](#footnote-17). Ironic Double Bluff is hence another case that is incorrectly ruled in by deceptionist accounts.

The counterexamples presented in this section may seem of secondary importance, but they issue a potentially devastating challenge to deceptionist accounts. A crucial selling point of the IDC is that it allows us to distinguish the believed-false statements that are lies from the believed-false statements that are not lies (the ones grouped under the label of non-assertoric falsehoods*)*. It turned out, however, that this distinction is inaccurate: some believed-false statements are both intended to deceive and not lies. This suggests that there must be some other feature that grounds the distinction between the believed-false statements that are lies and those that are not.

The most obvious alternative candidate is assertoric force. All non-assertoric falsehoods (metaphors, euphemisms, teasing remarks, etc.) are not lies and lack genuine assertoric force. In all counterexamples presented in this section, the statement does not count as a lie because it lacks the assertoric force that it would have if it were not ironic, fictional, jocular, etc. Hence, all believed-false statements that are not lies considered so far lack assertoric force, and can be excluded from the definition by requiring that the believed-false statement be genuinely asserted, thus avoiding the problems encountered by deceptionist accounts. This is the line of reasoning followed by assertion-based theories, that will be analysed in the following chapter.

If this observation is correct, the problem for deceptionist definitions becomes twofold. Not only do they fall victim to compelling counterexamples in every formulation; they also rest on a false assumption – that believed-false statements that do not count as lies can be singled out by appealing to their lack of deceptive intention. Counterexamples like Historical fiction and Ironic Double Bluff suggest that these cases are better singled out with a different criterion, namely lack of assertoric force. I will come back to this point in the next chapter; before examining assertion-based accounts, it is worth considering the third and most discussed family of objections to the IDC – namely, objections against the necessity of the condition.

## 4. Against necessity: lying without the intent to deceive

In the previous sections, we have witnessed the failure of the attempts to formulate the intent to deceive condition in a way that it is consistent with our intuitions, and the inability of deceptionist accounts to acknowledge that ‘deceptive non-assertoric falsehoods’ are not lies. These motives alone are strong enough to abandon the IDC and explore alternative conditions. But there are further compelling reasons to abandon the IDC: arguments against the necessity of the condition. Several philosophers (*e.g.* Carson 2006; Sorensen 2007; Fallis 2009) have argued that it is possible to lie without any intent to deceive, so that intending to deceive is not a necessary condition for lying.

The idea that lies are not necessarily aimed at deceiving has a particular appeal to contemporary philosophers, but is not a novelty in the philosophical literature. Classical authors like Augustine (*CM*, XIV), Peter Lombard (*Sen*, III, 38, 1; II, 213) and Aquinas (*ST*, II, q110, a2) included in their taxonomy of lies jocular lies, lies that are told “out of the desire to please” (Augustine, ibid.), “in order to make fun” (Aquinas, ibid.), rather than in order to deceive[[18]](#footnote-18), and several ancient philosophers defended a definition of lying in which the intent to deceive condition was not essential. This chapter will only focus on the contemporary philosophical literature.

### 4.1 Lying under duress

In the contemporary literature, Siegler (1966: 129) is the first author to present an example of lying without the intent to deceive:

Coerced Witness

“Suppose that A is forced by C's threat to tell *p* to B. A knows that *p* is false and that C intends that B be deceived into believing it. A hopes that B does not believe *p,* but he tells him and does nothing to prevent his believing *p.* Can we say that although (or because) A was forced to lie or to tell a lie to B, A did not intend to deceive B? If we can say this, then, intending to deceive is not a necessary condition for [lying].”

In the example, A is clearly lying to B, but A does not intend to deceive B – in any of the senses considered in the previous section. Thus, Coerced Witness represents a plain counterexample to any deceptionist account of lying.

Siegler (1966: 129) has pointed out that even if A (the speaker) has no intent to deceive B (the hearer), C (the coercer) has this intention, so that it might be argued that C satisfies the IDC. According to this interpretation, a third party’s intention to deceive the addressee is sufficient to make a believed-false utterance a lie (one might add the reasonable proviso: as long as the speaker is aware of that intention).

This interpretation seems ad hoc, and is easily rebutted by proving that we can easily elaborate versions of Coerced Witness where there clearly is no such third-party intention and the speaker is still lying. We could imagine that C (the third party) is innocent and believes A’s testimony to be true, even if A does not. In this case, in forcing A to tell what he believes to be false to B, C cannot have the intent to deceive B (because C believes the testimony to be veracious), but A would still be lying.

Now, in both cases A foresees that B might be deceived, and does nothing to prevent it (since otherwise C would hurt him). One could argue that, given that A is aware that deception is a probable consequence of his utterance, it is problematic to hold that A has no intent to deceive (Siegler 1969: 129-30). But foreseeing a probable side effect of an action is far different from intending that action to cause that effect[[19]](#footnote-19). As a matter of fact, A hopes that his testimony will not deceive B (so that C will be convicted without holding A responsible for the verdict), therefore A has no intention to deceive B. We can conclude that it is not problematic to hold that, even if A is aware that a possible side-effect of his utterance could be that B will be deceived, A’s utterance is not intended to deceive B.

### 4.2 Bald-faced lies!

Coerced Witness offers a first counterexample to the necessity of the IDC. However, even if in this example the speaker has no intention to deceive, it is rationally possible for the speaker to have the intention to deceive. But this does not help to redeem deceptionist accounts from the counterexample, as it is no reason to doubt that the speaker *can* lack such intent in uttering his statement. Perhaps, however, an even stronger case can be made against deceptionist accounts by providing further counterexamples, and in particular some in which it is *not* rationally possible for the speaker to have an intention to deceive.

Sorensen (2007, 2010:7) has provided a label for this kind of example, that he dubs “bald-faced lies”[[20]](#footnote-20). A bald-faced lie is the assertion of a believed-false proposition *p* in a context in which it is common knowledge[[21]](#footnote-21) that is epistemically impossible that the hearer believes *p* – hence, a context in which is impossible for the speaker to have the rational intention to deceive his addressee about *p.* More formally:

**Bald-faced lie**

An utterance with content p is a bald-faced lie iff:

1. S states p
2. S believes ¬p
3. “S believes ¬p” is common knowledge

(Given (c), S cannot intend to deceive A about p)

Claiming that bald-faced lies are genuine lies is equivalent to denying the necessity of the IDC, since in bald-faced lying the IDC is blatantly not satisfied. In the philosophical literature on the definition of lying, the first example of bald-faced lying tracks back to Carson (2006: 289-290), and is a variation on Coerced Witness, in which it is common knowledge that the witness is speaking falsely

Witness On Cctv

“Suppose that I know that the crime and my presence at the scene of the crime were recorded on a video camera so that there is almost no chance that the jury will believe that I believe what I am saying. Further, suppose that I am confident that I will not be charged with perjury, even if everyone believes that I am lying, […]. Given all of this [...], I lie by making a false statement, even though I have no intention of deceiving anyone about anything”

Witness On Cctv is a counterexample to the IDC in which the intent to deceive plays no role and cannot play any role, because the witness cannot rationally expect his testimony to be believed, given the CCTV footage.

Some might be sceptical about examples involving propositions uttered in an official context like a courtroom. It could be argued that utterances of this kind are not assertions, but speech acts of some other kind, like declarations (cf. III.3-4), that are merely meant to “go on the record” in the acts of the trial. However, there are plenty of other examples that avoids this objection[[22]](#footnote-22). A popular one is the cheating student example (Carson 2006: 21):

Cheating Student

“Suppose that a college Dean is cowed whenever he fears that someone might threaten a law suit and has a firm, but unofficial, policy of never upholding a professor’s charge that a student cheated on an exam unless the student confesses in writing to having cheated. The Dean is very cynical about this and believes that students are guilty whenever they are charged. A student is caught in the act of cheating on an exam by copying from a crib sheet. The professor fails the student for the course and the student appeals the professor’s decision to the Dean who has the ultimate authority to assign the grade. The student is privy to information about the Dean’s de facto policy and, when called before the Dean, he affirms that he didn’t cheat on the exam. He claims that he was not copying from the crib sheet. He claims that he inadvertently forgot to put his ‘‘review sheet’’ away when the exam began and that he never looked at it during the exam. The student […] warrants the truth of statements he knows to be false. He intends to avoid punishment by doing this. He may have no intention of deceiving the Dean that he didn’t cheat”.

In this example, the student states that he has not cheated, even if he is fully aware that the dean will not believe that he has not cheated. Even if he does not intend to deceive the Dean, intuitively he is lying.

## 5. Deceptionist Replies

When my love sweares that she is made of truth,

I do beleeve her though I know she lyes,  
That she might thinke me some untuterd youth,  
Unlearned in the worlds false subtilties.  
Thus vainely thinking that she thinkes me young,  
Although she knowes my dayes are past the best,  
Simply I credit her false speaking tongue, On both  
sides thus is simple truth supprest:  
But wherefore sayes she not she is unjust? And  
wherefore say not I that I am old?  
O loves best habit is in seeming trust,  
And age in love, loves not t'have yeares told.  
Therefore I lye with her, and she with me, And in  
our faults by lyes we flattered be.

William Shakespeare, Sonnet 138

### 5.1. Are bald-faced lies intended to deceive?

Partisans of the IDC have offered several responses against arguments based on bald-faced lies. A first way to dismiss these counterexamples is to argue that bald-faced lies are indeed intended to deceive, so that they are correctly captured by the IDC.

Meibauer (2016) follows this strategy, and alleges that it reflects ordinary intuitions about bald-faced lies: in other words, he claims that it is a shared intuition that bald-faced lies are intended to deceive. To support his claim, he conducted an experimental study on 128 German students. Participants were presented with a scenario involving a bald-faced lie (stories like Cheating Student or Coerced Witness) and asked to rate whether the target utterance was a lie and whether it was deceptive, on a 7-point Likert scale. All target utterances were considered to be both lies and deceptive. Meibauer (2016: 266) concludes that “this poses a major problem for the non-deceptionists”, given that it redeems the IDC from the most challenging of all counterexamples, namely bald-faced lies.

However, the data collected by Meibauer does not support his conclusion. The experiment only tested whether participants found the utterance *deceptive*. While *non-deceptionist* claim that bald-faced lies are not intended to deceive, they would hardly deny that bald-faced lies are, at least in some sense, deceptive. Because of the difference between being *deceptive* and being *intended to deceive*, the experiments fail to prove that laypeople have the intuition that bald-faced lies are intended to deceive. To dispel any doubts, Rutschmann and Wiegmann (2017) have conducted an experiment to verify if bald-faced lies are judged to be *intended to deceive*. Their data shows that, in some scenarios (as Coerced Witness), participants do ascribe *some* intention to deceive to bald-faced liars. However, they also found that in other scenarios people will deny that the bald-faced liar has any intention to deceive – especially when the speaker is *indifferent* to whether the audience will be deceived or not. While these results challenge the assumption that *no* bald-faced lies are judged to be intended to deceive, they also demonstrate that *some* definitely are – and this is enough to demonstrate that *deceptionist* definitions do not reflect laypeople’s intuitions about the concept of lying.

Rather than an empirical argument, to prove that bald-faced lies are intended to deceive Lackey (2013: 241-2) has tried to use a conceptual one. Simply put, her strategy is to show that the IDC can be broadened to capture bald-faced lies. More specifically, she suggests that aiming to *conceal information* counts as intending to *be deceptive*, and that an intent to conceal information is characteristic of bald-faced lies. Her revised IDC can be summarised as follows (Lackey 2013: 241):

**Intending to be deceptive condition**

(IDC£) S intends his statement to make A believe that p is true, or to conceal information from A as to whether p

To show how this is meant help with bald-faced lies, consider a very simple example. Imagine a scenario in which you leave a child alone in a room with a jar of jam – a room to which nobody else has access. Upon your return, the jam is empty, but the child insists on denying that he ate the jam. Given the circumstances, he might be aware that, no matter how hard he tries to convince you, you will not believe him. In this case, Lackey would say that he has no intention to make you believe that he did not eat the jam, but he still has the intention to conceal information from you as to whether he ate the jam. According to IDC£, the child’s bald-faced lie counts as *intended to be deceptive.*

There are several problems with this line of reply. First, it is far from obvious that concealing information is a form of deception (Mahon 2007: 187), and hence that capturing bald-faced lies with IDC£ would prove that bald-faced lies are intended to deceive. Second, the IDC£ is not a plausible alternative condition, as it falls victim to several of the counterexamples presented in the previous sections. Namely, it incorrectly excludes some *simple bluffs* (like Bandits) even if they are lies,and it incorrectly captures some *non-assertoric falsehoods* (like Historical fiction and Ironic Double Bluff), even though they are not lies.

Third, it is not clear if IDC£ succeeds in capturing all bald-faced lies as *intended to be deceptive* (cf. Fallis 2015). This is because you cannot succeed in concealing information from someone that already has that information (and thus cannot rationally intend to do so). This is clear in the Cheating Student example. We know that the dean already possesses the information, so it would be irrational for the student to intend to conceal that information. This is not to say that it is impossible for the student to have such intent; but since it is both possible and natural that the student does not have such intent, there is at least a version of Cheating Student that is a counterexample to IDC£. Similar observations apply to the jam scenario – since you know that it is impossible that the child did not eat the jam, it is just not clear which information he could be aiming to withhold by denying that he ate the jam.

Fourth, to prove that all lies are deceptive one has to demonstrate that it is impossible that a liar have no intent to conceal information (or to deceive), rather than demonstrate that *in a specific interpretation of a scenario* the liar has such an intent. No argument pointing to the former conclusion is found in Lackey (2013). Now, since (regardless of whether Fallis is right and in Cheating Student the student intended to conceal information) it seems conceivable that the student had no intention to conceal that he has copied (and that the Witness on CCTV had it, and so forth) by itself IDC£ does not pose a threat to counterexamples to the IDC based on bald-faced lying – at best, it may succeed in proving that some bald-faced lies are deceptive.

### 5.2 Are bald-faced lies assertions?

An alternative strategy to defend deceptionism is to argue that bald-faced lies are not lies, and therefore not counterexamples to the view that lying requires intending to deceive. Some authors simply bite the bullet in this respect, and plainly deny the intuition that bald-faced lies are lies (Mahon 2011, Meibauer 2011:282–83, 2014:108-9, Faulkner 2013). This would leave little space for counterarguments (even though some still stand, such as the ones explored in section 2 and 3), but there is certainly room to challenge the plausibility of this move. For instance, every empirical study conducted so far has proven that both ordinary speakers (Meibauer 2016, Rutschmann, R., & Wiegmann, A. 2017, cf. also Coleman and Kay 1981) and philosophy students (Arico & Fallis 2013) overwhelmingly agree that bald-faced lies are lies. This consensus is reflected in the literature: the vast majority of scholars writing on lying agrees that bald-faced are lies. While this is no conclusive proof that the deceptionists who share the opposite intuition are wrong, it is certainly a solid enough base to establish that their intuitions are non-standard, and that the onus is on them to offer strong reasons to think that the majority of philosophers and laypeople have incorrect intuitions.

In what follows, I will review some philosophical arguments that purport to offer such reasons. Broadly, all these arguments employ the same strategy: they all rely on the acceptance of (A), a view for which there is growing consensus in the literature:

1. Lying requires asserting a proposition

If one accepts (A), a new way opens up to demonstrate that bald-faced lies are not lies: if one can prove that bald-faced lies are not assertions, it follows that they are not lies. Such is the strategy pursued by these philosophers (Watson 2004: 72; Mahon 20009; Dynel 2011: 151; Leland 2015; Keiser 2015)[[23]](#footnote-23).

Before introducing any particular version of this argument, I must stress that this line of reasoning is unsound. It relies on the assumption that if there is clash between the intuition that (A) is true and the intuition that bald-faced lies are lies, it is the latter intuition that needs to be abandoned. This assumption is disputable. There is consensus that the *explananda* of a definition of lying are the *pretheoretical intuitions* about whether particular utterances constitute instances of lying: *theory-laden considerations* about the properties that are essential to lying can serve as a guide, but cannot be used as a litmus test to favour one theory over the other. Since (A) is a theory-laden consideration (it is a theoretical commitment of those who believe that asserting is a necessary condition for lying), it cannot be used to challenge pre-theoretical intuitions (*e.g.,* the intuition that a given bald-faced lie is a lie). Whenever there is a clash between the two, our methodological standards recommend that, absent other pre-theoretical intuitions pointing conclusively in the opposite direction (*i.e.* absent counterexamples to the negation of (A)), it is (A) that needs to be abandoned.

The upshot of these remarks is that any argument showing that some utterances that we would intuitively call lies (in this case, bald-faced lies) are not assertions would only licence the conclusion that (A) is false. Even if the authors that I am going to discuss do not interpret their arguments this way, also thus revised their argument yields a potential to defend *deceptionist accounts* of lying. I have mentioned that, at the current state of the art, the only alternative to deceptionism is represented by assertion-based definitions of lying. This is because only assertion-based definitions offer an alternative way to amend the classic definition in such a way that it excludes *non-assertoric falsehoods.* Quite obviously, assertion-based definitions of lying rely on the assumption that (A) is true. If the deceptionist can prove that (A) is false, assertion-based definitions are not an available alternative. While this would not redeem deceptionist definitions from counterexamples based on bald-faced lies, it would mean that they are the only account on the market that is able to distinguish between lies and irony, jokes, fictional utterances and the like. As such, deceptionist accounts would still hold way more explanatory power than the assertion-based alternatives (getting their predictions wrong for a much smaller subset of utterances), thereby still representing the best alternative on the market. In what follows I will review attempts to prove that bald-faced lies are not assertions, and show that they fail to undermine the project of defining lying in terms of assertions.

#### 5.2.1 Leland: Coercion Lies

In a recent paper, Leland argues that “some of the most prominent” (2015:1) examples of bald-faced lies in the literature are not lies, because they are not assertions. Leland attacks two counterexamples in particular, Witness on CCTV (from section 4) and Totalitarian State (below, from Seierstad 2003, reported in Sorensen 2007).

Totalitarian State:

“After the war starts, Asne Seierstad sneaks into a civilian hospital. She is surprised to see a ward with wounded soldiers. This suggests that Iraqi military hospitals are already overcrowded. – How may [sic] soldiers have you admitted today? I ask a doctor. – There are no soldiers here, the doctor says. – But they are wearing uniforms? – I see no uniforms, he says, and pushes me out. – You must go now, do you hear?”

As Leland correctly points out, both Witness on CCTV and Totalitarian State involve statements uttered because the speaker is forced to do so, rather than because he wants to do so. In the case of the Witness on CCTV, the witness is under a threat, and he can only choose to tell a lie or die. In Totalitarian State, the threat is more indirect; yet, the speaker here has no intrinsic motivation to lie, and choses to do so only because of external pressure. Leland’s argument is thus concerned specifically with lies uttered under duress, or *coercion-lies.*

Leland begins by defining asserting as undertaking a certain kind of responsibility, namely the responsibility of justifying your statement. Then, he argues that the statements uttered in Witness on CCTV and Totalitarian State are not assertions under this definition, and therefore not lies. Whether one accepts Leland’s definition of assertion or not, his worry that coercion-lies are not genuine assertions is certainly plausible, and shared by other scholars. For instance, Kenyon argues that “in a wide class of cases, duress seems to eliminate [...] the force of assertion [...] turning the utterance into a semantically structureless act of capitulation” (2003: 245). Similarly, for Watson (2003: 72) statements uttered under duress are not genuine assertions, and hence not lies.

A known reply to these arguments was formulated by Sorensen (2007: 257), who claims that if saying *p* under duress had no assertoric force, it would be equivalent to mentioning *p*. For instance, if a religious person were to be forced to utter “My God is a pig”, for his utterance to be non-assertoric it would need to be as void of assertoric force as its occurrence in the deliberate utterance of “I do not know how to say ‘My God is a pig’ in Arabic”. Nevertheless, it seems that there is a difference between the two cases: a religious person would be reluctant to proffer “My God is a pig” in the first case, but not in the second. This reluctance, Sorensen argues, is evidence that the coerced utterance is not void of assertoric force as Kenyon contends.

Sorensen’s counterargument is ingenious, but not entirely convincing. On the one hand, reluctance might be explained by the desire not to capitulate to the threat, thereby signalling submission to the force of the coercer (Kenyon 2010:364). On the other hand, it is not clear that an utterance lacking assertoric force needs to be on the same level as mentioning. Perhaps in uttering “God is a pig” the religious is doing something more than mentioning the proposition (he is repeating *p,* his repetition counting as accepting to follow an order) but also something less than asserting it (given the context, the utterance does not yield the commitments of a deliberate assertion). A tenable view is that saying something out of coercion comes with a greater commitment than merely mentioning it, and with less commitment than genuinely asserting it.

Finally, it seems that not all cases of coercions are alike. In some contexts, like in the case of the religious man being forced to utter blasphemies, it is common knowledge that the speaker is coerced. I share the intuition that in these cases some important component of assertion is amiss. In other cases, like Witness on CCTV or Coerced Witness, the audience is unaware that the speaker is coerced. Here it is not obvious that the speaker is not asserting. Since Leland defines asserting as taking responsibility for a proposition being true, it seems that in these contexts the speaker is indeed asserting: after all, he is claiming that something is true under oath, in front of people who are unaware of the coercion. As a result, also on Leland’s view only small subset of coercion-lies are not asserted.

This leads us to the second problem with Leland’s argument; even if successful, it would not be significant for the debate on lying. Contrary to Leland’s claim that Witness on CCTV and Totalitarian State are “two of the most prominent examples of bald-faced lies in the literature”, these two cases are instead the *least* discussed in the debate[[24]](#footnote-24), making his argument against them (that only seems to work for one of the two) not very consequential. More prominent and less controversial examples of bald-faced lies still stand, such as the cheating student example, so that the view that at least most bald-faced lies are assertions remains unchallenged.

#### 5.2.2. Implicatures

Leland’s arguments are ineffective because (amongst other reasons) they only attack a small subset of bald-faced lies. A similar weakness is found in arguments contending that bald-faced lies are not really asserted because they imply something else. Such arguments have the following structure: bald-faced lies require that it is common knowledge that what is said is false; by stating *p* when it is common knowledge that *p* is false, one necessarily ends up communicating something other than *p*, thereby failing to assert that *p* (Mahon 2009, Dynel 2011:151). This argument is clearly a valid reply in a subset of cases, like the following anecdote reported as a bald-faced lie by Sorensen (2007: 253, taken from Fadiman, 1985:481):

Maestro

“Once when the pianist Anton Rubinstein was practicing, the telephone rang inconveniently. His servant François picked up the phone and reported to the female caller that the maestro was not home. She objected “But I hear him playing”. “You are mistaken, Madame” insisted François “I’m dusting the piano keys” François loyally sustains the pretext that the maestro is not home. However, he is not being deceptive because he does not intend the caller to be taken in by such a preposterous lie”.

Here, it seems that François is not uttering a genuine assertion, because he is saying something that is blatantly false to politely imply something else – namely, that he is not going to disturb Rubinstein by passing him the phone (Mahon 2009, Faulkner 2013). François is using an expression that is blatantly untrue to avoid using another expression that would sound impolite. Arguably, Maestro is neither a case of lying nor of assertion, and consequently not a valid counterexample to the IDC (cf. Isenberg 1964:473, Mahon 2015:2.4).

While this argument is effective, it succeeds only because cases like Maestro are not genuine examples of bald-faced lies. In genuine cases, bald-faced liars do not imply the opposite of what they say, and they are genuinely asserting the proposition. For instance, the Cheating Student is not implying that what he says is false, as doing so would prevent his utterance from having the effect he wants – namely, to avoid the sanction. Sorensen’s choice of including cases like Maestro under the label of bald-faced lies is mistaken, and has generated unnecessary confusion. However, objections to this sort of examples will not generalise to other cases, and therefore will fail to prove that bald-faced lies are not asserted.

#### 5.2.3 Keiser: language games vs conversations

All the arguments reviewed so far fail at least partly because they cannot generalise to all bald-faced lies. Keiser (2015) offers an argument that avoids this accusation. Her key contention is that bald-faced lies are “instances of a move made in a language game without being moves in a conversation” (2015:417). Following Wittgenstein (PI), Keiser uses the expression “language games” to refer to “the myriad and varied activities in which we employ language” (2015:467). Conversations, by contrast, are that subset of language games that involves the performance of illocutionary acts, especially assertions.

Keiser claims that bald-faced lies are not moves in a conversation because they are not assertions[[25]](#footnote-25). More specifically, bald-faced lies lack the intention that would be required for them to qualify as assertions. The key claim here is that an utterance is an assertion *iff* the speaker R-intends[[26]](#footnote-26) that his saying that *p* provides the audience with a reason to believe that he, the speaker, believes that *p*. But in bald-faced lies, it is common knowledge that the speaker does not believe that *p,* so that the speaker cannot rationally have such intention. The conclusion is that bald-faced lies are not assertions, and therefore not moves in a conversation.

It is the first conjunct of the conclusion that is of interest here – that bald-faced lies are not assertions. The argument provided by Keiser is compelling; unfortunately, it only works if one accepts said definition of assertion. Given that it is possible to define assertion in some other way, her argument cannot be taken, like the others, to indirectly challenge the project of defining lying in terms of assertion: it merely sets a constraint on such project, namely the constraint that assertion-based definition of lying must employ a different definition.

This is hardly a problem, given that most contemporary philosophers would prefer to employ a different (*i.e.* non-Gricean) definition of assertion[[27]](#footnote-27). Currently such definition, once popular (cf. Chisholm and Feehan 1977, Bach and Harnish 1979), is widely regarded as indefensible. The most apparent problem with it is that an inveterate liar could not assert anything, because he could not hope to be taken seriously when he speaks (cf. Mahon 2008). More generally, any speaker who knows to be regarded as unreliable about *p* could not make any assertions with respect to *p.* Finally, this account does not capture bluffs (cf. section 2) as assertions (nor as lies), as in bluffing the speaker intends that the audience does *not* take his utterance as a reason to believe that he, the speaker, believes that *p*. In sum, Keiser’s attempt to prove that bald-faced lies are not assertions is also unsuccessful.

## 6. Conclusions

At the end of this chapter, one should be convinced that there are strong reasons to believe that one can lie even if one has no intent to deceive, against what is claimed by defenders of deceptionist accounts. A first line of argument has shown that each existing phrasing of the IDC is vulnerable to counterexamples. A second has noted that the IDC is too broad, as it incorrectly captures statements with no assertoric force that are intended to deceive, calling into question the ability of deceptionist accounts of lying to treat *non-assertoric falsehoods* correctly. A third has shown that one can lie without intending to deceive the hearer, so that the IDC is not a necessary condition for lying. None of the replies has succeeded in challenging these arguments successfully.

All the arguments point to a single conclusion: the intention to deceive is not a necessary condition for the definition of lying, and the commonality between believed-false statements that are not lies (*non-assertoric falsehoods*) is not to be found in the absence of a deceptive intention.

A final worry might be that a ‘non-deceptionist’ definition of lying (*i.e.* any account requiring at least that the speaker utter an insincere statement, and not requiring that the statement be intended to deceive) would fail to identify the deceptive nature of lies, and to capture what is morally wrong with lying. Let me deal with each of these alleged problems separately.

The first worry, that non-deceptionist definitions would not capture the deceptive nature of lies, is misguided. First, to admit that it is possible to lie without the intent to deceive is not to claim that lying necessarily does not involve deception, nor to deny that most lies are aimed at deceiving and many succeed to deceive. In fact, philosophers who reject the IDC acknowledge that the most common cases of lying involve an intention to deceive, and concede that even lies with no intent to deceive are likely, in some cases (like Coerced Witness), to deceive the addressee (Carson 1988, Sorensen 2007, Fallis 2009; 2010).

Second, when philosophers like Lackey (2013) complain about an “unhappy divorce” between deception and lying, it should be clarified that deception and lying have never been married, since neither deceptionist nor non-deceptionist accounts require successful deception as a necessary condition. If the claim is that deception is intuitively connected with lying, both deceptionist and non-deceptionist accounts allow for its absence. If the claim is instead that the intent to deceive is intuitively connected with lying, then we have already seen that there are compelling reasons to think that it is frequent, but not necessary, that one lies with the intent to deceive – so that there is a connection, but not an essential one.

For similar reasons, the worry that a non-deceptionist definition of lying would necessarily fail to capture what is morally wrong with lying is misguided – and this independently of the ethics of lying that one is keen to endorse. The reason is that non-deceptionist definitions still require the speaker to say the opposite of what he believes, and this amounts to a violation of the norm of truthfulness, a violation that is able to explain the prima facie wrongness of lying (this is, for instance, the received Augustinian and Kantian view). Moreover, both non-deceptionist and deceptive lies are attempts to manipulate people, and arguably manipulating people is morally wrong – whether or not this is obtained by deceiving (Bok 1978:18-20; Williams 2002:93; Fallis 2010; 2014).

In conclusion, given that deceptionist are not able to provide sufficiently convincing reply to the diverse counterexample to which they fall victim, and noted that abandoning the IDC does not yield problematic consequences as alleged, we can happily abandon the IDC. This raises the question of what condition can replace the IDC. The next chapter will propose a solution: to require that the speaker genuinely asserts the propositional content of his statement.

# III. Lying: a speech-act theoretic approach

Hamlet: *I think it be thine [the grave],* *indeed; for thou liest in't.*  
First Clown*: You lie out on't, sir, and* *therefore it is not yours: for my part, I do* *not lie in't, and yet it is mine.*  
Hamlet: *Thou dost lie in't, to be in't and* *say it is thine: 'tis for the dead, not for the* *quick; therefore thou liest.*  
First Clown: *'Tis a quick lie, sir; 'twill* *away gain, from me to you.*

Shakespeare, *Hamlet*, V, 1

Let us start with a quick reminder. We have seen that almost every author writing on lying agrees on the following *necessary* conditions for defining lying (previously introduced as the *Classic Definition).*

**Necessary conditions for lying**

S lies to A only if:

(a) S says p

(b) S believes ¬p

The contemporary debate revolves around the question of how to supplement this account with further necessary conditions, so that altogether the definition will provide jointly sufficient conditions for an utterance to be a lie. More specifically, the extra condition(s) will need to rule out *non-assertoric falsehoods:* statements that are believed to be false but that are not lies.

Scholars are divided depending on whether or not they accept the intention to deceive (condition (c) below) as an extra condition. Authors who adopt this condition are called *deceptionist;* authors reject it are call *non-deceptionist*. Typically, non-deceptionist philosophers employ *assertion-based definitions,* requiring instead that the speaker must *genuinely assert* what is said (condition (c’) below).

**Deceptionist definitions:**

S lies to A iff: (a) S says p, (b) S believes ¬p, and (c) S intends A to believe p

**Assertion-based definitions:**

S lies to A iff: (a) S says p, (b) S believes ¬p, and (c’) S asserts that p

In the previous chapter, I have presented conclusive reasons to think that deceptionist definitions do not offer an analysis of lying that matches our intuitions. This chapter considers whether assertion-based definitions offer a viable alternative.

Since assertion-based definitions differ depending on how the ‘assertion-condition’ (c’) is formulated, in section 1 I introduce its most influential versions. I argue that each of the existing accounts falls victim to counterexamples, and that where previous accounts have failed, an account based on a speech-act theoretic notion of assertion can succeed. To strengthen my case, in section 2 I consider three speech-act theoretic puzzles that present a challenge for all the existing accounts (both deceptionist and assertion-based ones). Drawing on some insights from speech act theory, section 3 develops an alternative assertion-based definition, that avoids the difficulties encountered by previous definitions. In section 4, I show that my proposed definition avoids all the difficulties that affect other accounts, and that it has the resources to solve the speech-act theoretic puzzles introduced in section 2.

## 1. Assertion-based definitions, and their difficulties

Authors endorsing an assertion-based definition of lying tend to agree that a speaker lies iff *he asserts something insincerely*[[28]](#footnote-28), but disagree on what counts as an assertion for the purpose of such definition. The three most prominent accounts of the assertion-condition are given by Carson (2006, 2010), Fallis (2009, 2012, 2013) and Stokke (2013)[[29]](#footnote-29). Respectively, for these authors the ‘assertion-condition’ (c’) is:

ACC: S intends to warrant the truth of p (Carson 2010: 37)

ACF: S believes that he states p in a context where the following norm of conversation is in effect: “Do not make statements that you believe to be false” (Fallis 2009)

ACS: S proposes that p become part of the official common ground (Stokke 2013)

In what follows, I will present each of these proposals more in detail, and consider whether they provide a good condition for the purpose of defining lying. As a terminological clarification, whenever I will refer to an author’s *definition of lying* as opposed to their assertion-condition, I will mean the definition of lying that results from plugging the assertion-condition into the ‘assertion-based definition’ schema (a-b-c’).

### 1.1. Carson

ACC: S intends to warrant the truth of p

In developing his proposal, Carson follows a suggestion by Ross (1930), according to whom lying is breaking an implicit promise to tell the truth (also Hartman 1975: §II, 286, Fried 1978:67). ACS is meant to reflect this idea: “if one warrants the truth of a statement, then one promises or guarantees, either explicitly or implicitly, that what one says is true” (Carson 2006: 294). Lying is therefore understood as warranting the truth of something you do not believe to be true.

A known difficulty for this account is represented by *proviso-lies* (Fallis 2009, Arico & Fallis 2013): lies where the speaker makes it explicit that he is not guaranteeing that what he says is true. Here is a (slightly revised) example of proviso-lie from Arico & Fallis (2013):

Wild Party (proviso-lie)

Last night, after a particularly wild party, Chris found her swimming trophy broken. Today Chris is trying to figure out who broke her trophy. Chris says to Jamie, "So, somebody was in my room last night and broke my trophy. Did you see anything?". Jamie clearly remembers that she was the one who broke Chris's trophy. Since everyone knows that Mel is always breaking stuff, Jamie responds to Chris:

1. Yeah, um, Mel broke your trophy.

(1\*) But I was kinda drunk, and there were lots of people in there, so don’t take my word for it.

In this example, Jamie’s statement (1) is followed by a ‘proviso’ (1\*) that is meant to clarify to that Jamie is not guaranteeing to Chris that (1) is true. Since, as a result, Jamie’s does not warrant (1\*) as true, ACC does not count (1) as a lie. Nonetheless, Jamie is clearly lying, so that proviso-lies like (1) are a counterexample to Carson’s definition of lying.

Carson (2010:36-9) replies that since “warranting comes in degrees of strength, a moderately strong assurance of truth is all that is required for lying”. Since (1) comes with a moderate assurance of truth, it can after all be dubbed a lie on his account. The problem with this reply is that the claim that “warranting comes in degrees of strength” is not really compatible with Cason’s account of warrant. On his view, warrant is basically an implicit promise, and promises do not come in degrees. Promises cannot be mitigated or downgraded, and there is no sense in which they can give a “moderately strong” assurance of truth: either they warrant that something will be the case or they do not (cf. Fallis 2013:347-8). To see this, consider the difference between adding a proviso to an assertion and adding a proviso to a promise:

(2a) I am willing to come to the cinema tomorrow, but you know that I am really bad with dates and times

(2b) # I promise that I will come to the cinema tomorrow, but you know I am really bad with dates and times

While (2a) is a mitigated assertion, (2b) is not a mitigated promise: it is not a promise at all. Promises do not come in degrees, and promises followed by the kind of proviso involved in proviso-lies are simply not promises. Perhaps the notion of warranting can be amended to avoid the link with promises (I follow a similar strategy in section 3); but as long as the notion of ‘warrant’ is conceived in analogy with promises, ACS will be unable to deal with *proviso-lies*.

### 1.2. Fallis

ACF: S believes that he states p in a context where the following norm of conversation is in effect: “Do not make statements that you believe to be false”

Fallis' proposal draws on the theory of cooperative communication developed by H.P. Grice (1989), according to which cooperative speakers are expected abide by several conversational maxims, one of which requires them not to say what they believe to be false. Fallis’ proposal is to understand lying as the deliberate violation of this sincerity maxim, in contexts in which the speaker is expected to follow the maxim (as opposed to contexts in which the maxim is suspended as, *e.g*., on a theatre stage). A known problem with Fallis’ definition is that it counts most ironical utterances as lies (Stokke 2013). To see this, suppose that, on an extremely stormy day, I ironically utter (3) while I am ordering a falafel wrap:

(3) What nice weather outside!

In stating (3) ironically, I say what I believe to be false, but I am clearly not lying. However, according to Fallis' definition (3) is a lie, because I say what I believe to be false in a context (a casual conversation while ordering a falafel) in which the norm of sincerity is not suspended (cf. Stokke 2013 for further discussion).

Fallis has attempted to revise his definition in order to avoid these counterexamples. A first alternative proposal is the following (Fallis 2012:577):

DLF: you lie iff you intend to communicate something false by saying that thing.

The acronym DLF here stands for ‘Definition of Lying by Fallis’, to stress that this clause is not just an assertion condition, but a complete definition of lying. Fallis’ assertion condition could be extracted from DLF as:

ACF2: S intends to communicate that p

DLF successfully rules out cases of irony: when I utter (3), I state what I believe to be false, but I do not intend to communicate that there is nice weather outside – rather the opposite, namely that the weather is bad. However, whilst this proposal avoids the problems of the earlier version, it is still unsatisfying. Its main issue is that it relies on an unspecified notion of ‘communication’. Fallis acknowledges that the notion plays a key role in DLF: “what counts as communication makes a difference for what counts as a lie [according to DLF]”, and nonetheless controversially (and unconvincingly) adds that no notion of communication is needed for his account to work: “for purposes of this paper, it will not be necessary to settle on one specific account of communication (2012:572)”.

To be sure, Fallis does make some suggestions. He specifies (reasonably so) that he does not mean communication in a Gricean sense, as that would require the speaker to intend to make the audience believe that *p*[[30]](#footnote-30), reintroducing the intention to deceive condition. He mentions a couple of accounts of communication that would avoid this difficulty, but does not in fact pick one. As long as the notion of communication is left unspecified, DLF will be underdetermined. And absent clear predictions, it is not possible to argue that a given example contradicts DLF’s predictions, so that it is not clear how DLF can be falsified (see also Keiser 2015:476,fn26).

An underspecified and unfalsifiable definition must certainly be discarded, but it could be argued that there is at least one reading of DLF that makes defined predictions. Fallis might not have specified a technical notion of communication simply because he has in mind a rather ordinary notion of communication. On this reading, what counts as ‘communicating’ for the purpose of DLF is what an ordinary speaker would regard as ‘communicating’. This gives us a criterion to test the definition, and to produce a putative counterexample. Consider the following scenario (for a similar one, cf. the Historical Novel scenario in II.2.3):

Reptilian President

Gunther is a well-known writer of several works of fiction, that span from spy stories to erotic novels. Gunther found out that about 12 million Americans believe that reptilians are among us[[31]](#footnote-31). He thinks that the majority of reptilian-believers voted for Donald Trump, and that the best way to change their electoral opinions is to convince them that Donald Trump is a reptilian. To this end, he writes a fictional spy story set in our days, in which the protagonist discovers many shocking and unfalsifiable details about Trump's identity as a reptilian, hoping that the gullible conspiracy theorists that constitute his audience will interpret the novel as containing some truthful revelations about the reptilian plan to take over the US. His fictional novel contains several statements like (4):

(4) Donald Trump is a reptilian

Now, in his novel Gunther states that (4) and believes that (4) is false. Intuitively (according to an ordinary notion of communication), he intends to communicate that (4) to his conspiracy-theorist audience: that is why he has written the novel in the first place. Nonetheless, against DLF's prediction, Gunther's statement that (4) is not a lie, because it is a boldly implausible statement contained in a work of fiction, and we do not regard boldly implausible statements contained in works of fiction as lies.

We can conclude that, on a standard understanding of communication, DLF is too broad. Surely there are ways to revise the definition to avoid counterexamples of this kind. Absent such revisions, however, DLF is both underspecified, and too broad if "communicate" is understood in a commonsensical way.

In a more recent paper, Fallis has proposed a more promising definition of lying – arguably, one that represents a refinement of ACF2, *i.e.* one that specifies one sense in which we might understand the notion of ‘intending to communicate’. Drawing on some observations by Davidson (1985; 2001), Fallis (2013) proposes the following assertion-condition for defining lying:

ACF3: The speaker intends to represent himself (to his audience) as believing that p is true

To represent yourself as believing something is to present yourself as having a particular property, namely the property of believing a proposition. Fallis correctly points out, following Davidson, that we have an intuitive grasp of this notion, and this becomes evident when we think about familiar cases: in signing a cheque, you *present yourself* as having enough money in the bank to honour the cheque (Black 1952:31); in wearing a cross necklace, you *present yourself* as being Christian; in playing chess, you *present yourself* as trying to win, and so forth. Furthermore, the idea that assertions represent the speaker as believing a proposition is familiar to philosophers, because it bears striking similarities with the popular view that assertions are the outwards expression of a belief – a view that (under different guises[[32]](#footnote-32)) has been widely defended in the literature.

Unlike ACF, ACF3 does not count irony as lying: when you utter *p* ironically, you do not represent yourself as believing *p.* ACF3 is also an improvement over ACF2, since it gives us a clearer criterion to predict whether a given utterance is a lie. Furthermore, ACF3 correctly rules out standard cases of fiction: for instance, an actor reciting a line on stage does not represent himself as believing what he says. However, like ACF2, ACF3 has some difficulties in treating *insincere* fictional utterances. For instance, in Reptilian President Gunther intends to convince his readers that (4)is true, so that it makes sense for him to intend to represent himself as believing (4). This seems the best way to interpret the scenario: after all, he wrote the novel in order to present himself as possessing some inside knowledge of the reptilian conspiracy theory. If we read the scenario in this way, ACF3 incorrectly counts (4) as a lie.

One could rightfully observe that Gunther’s intention is somewhat unconventional: it is not normal for writers of sci-fi novels to try to present themselves as believing the stories they narrate. But even if we grant that Gunther’s intention is unusual, this does not mean that it is *irrational* for Gunther to have such intention – and while ACF3 should be interpreted as requiring a *rational* intention (cf. Fallis 2015), it does not seem to require a *conventional* one. Furthermore, in realistic fictions (as the Historical Novel from III.3) it is conventional for the author to represent himself as believing many of the propositions in the novel. Insincere realistic fictions are thus a clear counterexample to ACF3; insincere fictional utterances in general pose perhaps a less serious threat, but they are not obviously ruled out by this account.

### 1.3. Stokke

ACS: S proposes that p become part of the official common ground (Stokke 2013)

Like Fallis’ ACF, Stokke’s assertion-condition draws on an existing proposal: the account of assertion and the notion of *common ground* developed by Stalnaker (1978, 1998, 2002). According to Stalnaker (2002: 716), “it is common ground that *p* in a conversation if all members *accept* (for the purpose of the conversation) that *p*, and all *believe* that all accept that *p,* and all *believe* that all *believe* that all accept that *p*, etc.”.

Stokke acknowledges that defining assertion as an attempt to add *p* to the common ground of the conversation would not do the job[[33]](#footnote-33), because not all such proposal are assertions. For instance, if an interlocutor proposes to accept *p* for the sake of the argument, or as a hypothesis, he is proposing to add *p* to the common ground while not asserting that *p.*

In order to narrow down his definition, Stokke introduces a distinction between *official* and *unofficial* common grounds (henceforth CGs). Unofficial CGs are ‘temporary’ CGs that open up in order to store an information that is used for the purpose of an argument; by contrast, official ones are, so to say, ‘permanent’ CGs. ACS defines assertion in terms of *official* CG, so as to exclude the potential counterexamples considered earlier (assumptions, hypothesis, etc.).

Against this definition, Fallis (2013) has argued that it is possible to lie by asserting a believed-false proposition without proposing that it becomes part of the official CG. Fallis (2013: 23) clarifies this point by means of an example:

 Deep Throat

“A devious Deep Throat attempts to mislead a journalist by saying:

(5a) I am only saying this to you.

(5b) And I am only going to say it once.

(5c) If you repeat it (or say anything that presupposes it), I will deny it.

(5d) The Attorney General himself was behind the cover-up.

Deep Throat does not seem to be proposing [(5d)] be added to the common ground of the conversation. [...] Deep Throat is quite clearly proposing that this information not be added to the common ground. Even so, Deep Throat seems to be lying as he knows that the Attorney General was not behind the cover-up”.

This counterexample seems to disqualify Stokke’s account. However, Stokke (2017) presents three counterarguments aiming to show that (5d) is indeed added to the CG. Let us see them in turn.

(A) *Pretence:* Stokke claims that in uttering (5a-c), Deep Throat (henceforth DT) might be merely announcing that he *will pretend* that (5d) is not in the CG. Since pretending that *p* is not in the CG presupposes that *p* is in the CG, on this interpretation (5d) has indeed been added to the CG.

(B) *Presupposition*: in support of his first argument, Stokke points out that whenever a proposition can be felicitously presupposed, we can take it as good evidence that that proposition is indeed in the CG. As a matter of fact, (5d) can be felicitously presupposed: for instance, the journalist might felicitously reply:

1. How did the Attorney-General arrange the cover-up?

(C) *Denial:* Stokke points out that Fallis’ argument seems to rely on the assumption that (5c) (“If you repeat it (or say anything that presupposes it), I will deny it”) prevents (5d) from becoming the CG. However, while it is obvious that (5c) establishes some reasons not to repeat or presuppose (5d), it is not clear that it prevents it from becoming CG. A proposition can be in the CG even if there are external reasons that prevent participants in the conversation from repeating or presupposing that proposition, and (5c) only seems to establish such reasons. Other, more mundane examples of such reasons are considerations of politeness (it is not polite to repeat/presuppose that *p*), safety (it is not safe to do so), empathy, etc. Thus, while we can agree that (5c) establishes reasons not to repeat *p*, this does not prove that *p* is not in the CG.

Stokke's counterarguments are well constructed, but they miss the point. A first problem is that they rely on treating the notion of CG as if it was equivalent with the notion of common knowledge/belief[[34]](#footnote-34), despite the important differences between the two notions. Surely argument (A) works for common knowledge: in the example it is common *knowledge* that (5d), and arguably DT is merely announcing that he will pretend that (5d) is not in the common knowledge. But while this remark is valid for common knowledge, it does not work for CG. Remember that for *p* to be part of the CG, each and every membermust *accept that p for the purpose of the conversation* (rather than know it or believe it). In our example, DT does not agree to accept *p* for the purpose of the conversation, and makes it very explicit that he is not willing to do so (as clarified by his uttering (5a/c)). Therefore, against (A), (5d) is arguably common knowledge, but clearly not CG.

A similar reply applies to (B). To establish that (5d)is in the CG, Stokke points out that the journalist is able to felicitously presuppose that (5d) is true. However, by Stokke’s own admission (2017:179), felicitous presupposition of a proposition *p* is not conclusive evidence that *p* isin the CG. Arguably, one reason is that a proposition can be felicitously presupposed when it is common knowledge but not CG. Felicitous presupposition of (5d) can be explained by both assumptions; since we have already seen that the assumption that (5d) is CG does not fit the scenario, whereas its being common knowledge does, the best explanation for (5d) being felicitously presupposed is that it is common knowledge (but not CG) that (5d) is true.

Finally, a general worry with Stokke’s arguments is that they do not try to prove that DT has *proposed* to add (5d) to the CG; they merely try to prove that (5d) has *as a matter of fact* been added to the CG. This is a relevant difference, because a proposition can become CG independently of someone’s proposal to make it become CG. For instance, in virtue of my sneezing it may become CG in a conversation that I have just sneezed, even though I have made no proposal to add this proposition to the CG. And by Stokke’s own admission, whether you propose to add something to the CG is a matter of what you are trying to do, rather than a matter of your audience’s uptake (2013:59).

These observations help to defuse argument (C). What is at stake in the scenario is whether DT is *proposing* to the journalist to accept (5d) for the purpose of the conversation, not whether (5d) is as a matter of fact added to the CG. Stokke may be right that giving your interlocutor a reason not to repeat or presuppose *p* (uttering (5c)) is not always sufficient to rule *p* out of the CG. But what is at stake is what DT is trying to do, and DT is *ex hypothesis* not *trying* to add (5d) to the CG (he has the opposite intention). His utterance (5c) is meant to make this intention explicit: whether or not it succeeds in preventing (5d) from becoming CG is not relevant to ACS. Stokke’s argument (C) is thus unsuccessful, since it does not show that (5d) is captured by ACS. Since all three arguments fail, the conclusion is that (5d) cannot be counted as a lie according to Stokke’s proposed definition.

## 2. Three Puzzles From Speech Act Theory

At this point, it seems that all the assertion-based definitions on the market are inadequate to define lying. This becomes even more evident when we consider the phenomenon of lying within a speech-act theoretic perspective. In what follows, I will show that when speech acts other than lying are taken into consideration, new problems for all existing definitions emerge. In considering these problematic cases, I am setting up the ground for my proposed solution: to understand lying as a particular kind of assertive speech act.

### 2.1 Explicit Performatives

What is an “explicit performative utterance”? Broadly, an explicit performative utterance occurs when:

1. a sentence is uttered and an action is thereby performed, and
2. the grammatical form of the sentence makes it look at first glance as though the speaker states that he performs that action (Heal 1974:106)

To give an example, in uttering (1) the speaker is *assuring* the audience that (1\*) by stating that he is assuring that (1\*), and in uttering (2) he is *promising* that (2\*) by stating that he is promising that (2\*):

1. I assure you that (1\*) [I stole the cake]
2. I promise that (2\*) [I will clean the toilet today]

More generally, the standard form of an explicit performative utterance (henceforth, performative utterance for brevity) is (PU), and the verb that occupies the [performative verb] slot is called an *explicit performative*:

(PU) “I (hereby) [performative verb] that *Φ*”

Intuitively, performative utterances like (1) and (2) can be lies: if I utter (1) and I did not steal the cake (*i.e.* I believe (1\*) to be false), I am clearly lying; similarly, I am lying if I utter (2) and I intend not to clean the toilet today. Nonetheless, all definitions in the literature seem to counterintuitively predict that performative utterances cannot be lies (Mahon 2010:50’’).

To see this, recall that all definitions of lying (the *classic definition*, and both *deceptionist* and *assertion-based* ones) require that the speaker believes that *what he said* is false[[35]](#footnote-35). In other words, there is agreement that at least (a) and (b) are necessary conditions for lying:

**Necessary conditions for lying**

S lies to A only if:

(a) S says p

(b) S believes ¬p

Now, in uttering (1), what the speaker says (*p*) is (1)*:* “I assure you that I stole the cake”. Since in the very act of uttering (1) the speaker *makes it the case* that he assured his interlocutor that he stole the cake, a speaker cannot rationally believe *p* to be false, so that condition (b) cannot be satisfied. In other words, a rational speaker cannot lie in uttering (1), or in uttering any sentence following the PU schema. Searle (1989) talks in this sense of a ‘self-guaranteeing’ nature of performative utterances: “performative utterances are self-guaranteeing in the sense that the speaker cannot be lying, insincere, or mistaken about the type of act being performed”.

Although the verdict that performative utterance cannot be lies is counterintuitive, scholars who considered the issue are willing to bite the bullet in this respect. Meibauer (2014:76), for instance, accepts that insincere promises and warnings executed by means of a performative utterance cannot be lies. But this seems wrong: some performative utterances (including the previous examples) are *clearly* lies, and denying them this status has some counterintuitive consequences. One is that we will not be able to count some assertions as lies, even if most scholars (including Meibauer) agree that insincere assertions are lies. For instance, in uttering (3) I am clearly asserting that (3\*) (that Tony was not present at the scene of the murder), and my assertion is insincere if I believe (3\*) to be false:

1. I assert that (3\*) [Tony was present at the scene of the murder]

It seems that the problem of treating performative utterances with traditional definitions of lying is simply that these definitions of lying fail to identify which belief determines the speaker’s insincerity: they should acknowledge that insincerity is determined by believing that (3\*) is false, rather than believing that (3) is. Rather than following Meibauer in claiming that performative utterances cannot be lies, scholars could simply adopt a different semantics of explicit performatives, one that identifies (3\*) as the relevant proposition *p*.

The semantics of performative utterances adopted by traditional definitions is typically named *descriptivist*[[36]](#footnote-36)in the literature on performative utterances. *Non-descriptivist*[[37]](#footnote-37) accounts,by contrast, maintain that an utterance like (3) is uniquely a direct assertion with content (3\*), and that in uttering (3) the speaker does not assert (3). Non-descriptivist accounts allow for performative utterances to count as lies, because they identify the content of the assertion in a way that allows for insincerity: whenever (3\*) is believed to be false, (3) is a lie. Descriptivist accounts, by contrast, do not allow performative utterances to be lies: (3) cannot be a lie, because (3\*) is counted as implicated rather than asserted – so that a performative utterance can be misleading, but not lies.

If standard definitions want to capture some performative utterance as lies, they need to alter the statement condition (a). Instead of requiring that the speaker says *p*, they need to require that the speaker utters a sentence with content *p*, where *p* is identified in a non-descriptivist fashion. The revised necessary conditions for lying will read as follows:

**Revised necessary conditions for lying**

S lies to A only if:

(a’) S uttered a sentence with (non-descriptively identified) content p

(b) S believes ¬p

This revision will work for both performative utterances and non-performative utterances: in the former case, content differs from what is said; in the latter, the two coincide – but in both cases, *p* correctly identifies the proposition that, when believed to be false, could make the utterance count as a lie. This relatively simple modification allows performative utterances to count as lies on all pre-existing definitions. At the same time, however, it elicits new questions about performative utterances: can all performative utterances count as lies? If not, which ones can count as lies, and under which conditions?

### 2.2. Determining which speech acts can be lies

Once one considers the question, it seems that some performative utterances can count as lies, while others cannot. Compare the previous cases (repeated below as (1) and (2)) with some new cases:

1. I assure you that (1\*) I stole the cake
2. I promise that (2\*) I will clean the toilet today
3. I order you to clean the toilet (4\*) [that you will clean the toilet][[38]](#footnote-38)
4. Assume that (5\*) the defendant is guilty…
5. I bet that (6\*) Miklos committed the murder!

Intuitively, it seems that only the two original performative utterances (1)-(2) can count as lies. Clearly (4) cannot be a lie, as it is not even clear what could make (4) insincere: if I do not believe that you will clean the toilet (or I do not want you to clean the toilet), it is not clear that (4) would be insincere, much less that it would be a lie. Similarly, if I invite you to assume that the defendant is guilty for the sake of the argument, as in (5), it does not seem that I am lying, no matter how devious my argumentative strategy and my intentions are. Lastly, while I can mislead in uttering (6) when I know (6\*) to be false (*e.g.* I committed the murder myself), it does not seem that I could be accused of lying for betting something.

Let us assume that a good definition of lying should dub the insincere utterance of (1-2) as lies, and of (4-6) as not lies. Unfortunately, none of the theories that we have considered so far (except, perhaps, for Carson’s ACC) clearly makes these predictions. I will consider them in turn, to see to which extent their predictions diverge from our desiderata. To begin with, a revised deceptionist definition could read as follows:

**Deceptionist definition (revised for PU)**

S lies to A only if: (a’) S uttered a sentence with content p, (b) S believes ¬p (c) S intends A to believe p

According to this definition, it can be possible for (4-6) (see reminder in footnote below) [[39]](#footnote-39) to be lies, because it is possible that the speaker believes the content to be false and intends the hearer to believe that it is true. For instance, it is possible for a judge to believe that his assumption (5\*) is false, and to state (5) with the intention that the audience comes to believe that it is true. The standard definition thus fails to make the required predictions.

What about assertion-based definitions, that replace condition (c’) with an assertion-condition? If the assertion-condition is interpreted simply as requiring that the speaker asserted something (Sorensen 2007), none of these utterances can count as lies, not even (1) or (2): (1) is an assurance rather than an assertion, (2) is a promise, (4) an order, and so forth. However, none of the assertion-based definitions considered sets such a literal requirement. Rather, these accounts provide an assertion-condition that must be satisfied for the utterance to count as a lie. In principle, then, speech acts other than assertion might be captured as lies, as long as they satisfy the relevant ‘assertion-condition’. As a reminder, the four relevant versions of this condition are:

ACF2: S intends to communicate that p

ACF3: S intends to represent himself (to his audience) as believing that p is true

ACS: S proposes that p become part of the official common ground

ACC: S intends to warrant that p

It seems that ACF2 makes the same incorrect predictions of standard accounts: in each case considered, it is possible for the speaker to intend to communicate something that he believes to be false. For instance, suppose that I utter (4) and I believe that you will disobey my command. In this case, I am still attempting to communicate to you that you will clean the toilet (remember that ‘communicating’ is not understood as a success verb in ACF2, so that you communicate that *p* even if your audience does not come to believe that *p*), even if I believe that you will not clean the toilet.

ACF3 makes better predictions. It rules in assurances and promises (1-2) as lies, because they represent the speaker as believing the relevant proposition to be true. It also rules out assumptions like (5) [[40]](#footnote-40): in assuming that (5\*), the speaker may represent himself as *accepting* the proposition for the purpose of the conversation, but not as *believing* it. The verdict for commands is less clear. In many contexts, in uttering (4) a speaker will form the belief that the recipient of the order will make (4\*) true, so that it will be natural to ascribe him the belief that (4\*). In these cases, it might be argued that the speaker is representing himself as believing (4\*). However, it is not clear that ACF3 is committed to this prediction; here its verdict is less straightforward, but not obviously wrong.

The real problem for ACF3 are bets. In uttering (6) (see footnote), I can clearly represent myself as believing that (6\*): if I believe that (6\*) is false, ACF3 incorrectly predicts that my utterance is a lie. The problem is not limited to bets: a similar case against ACF3 can be made by considering guesses, predictions, conjectures, and other speech acts that cannot be lies, but express a belief.

One might reply that in making a bet like (6) you do not represent yourself as believing its content (6\*), but merely as having some confidence in its truth. Now, we could plausibly interpret ACF3 as requiring that the speaker expresses an *outright belief* in the relevant proposition. If (6) merely expresses a *partial belief* in the proposition, the revised ACF3 does not count these utterances as assertions, nor as lies (for more on the distinction between *partial* and *outright* belief, cf. IV.4.3).

There are two problems with this reply. The first (and most important) one is that ACF3 only requires that the speaker *intends* to represent himself as believing something. Surely (6) can be used to express a partial belief; but this does not mean that a rational speaker cannot utter (6) with the intention to express an outright belief – as a matter of fact, this appears to be a relatively standard use of (6). When uttered with the intention to express an outright belief, insincere bets like (6) are captured as lies even by the revised version ACF3.

The second problem is that ACF3 cannot afford to discard statements expressing partial beliefs, because statements expressing partial beliefs can clearly be lies. For example, consider the following statements:

(6a) Miklos has probably committed the murder

(6b) Miklos committed the murder. But I am very bad at recognising faces, and the room was very dark, so I am not sure of it

(6\*) Miklos committed the murder

In uttering (6a), the speaker expresses a partial belief in the truth of (6\*). Intuitively, if he believes that (6\*) is highly unlikely or impossible, (6a) is a lie. If we restrict ACF3 to utterances expressing outright beliefs, we have the undesired result that these utterances are lies[[41]](#footnote-41). Similarly, the *proviso-lie* (6a) expresses a partial belief in (6\*). Unless we allow for utterances expressing partial beliefs to count as lies, we will have that proviso-lies of this sort are also not genuine lies – a prediction that Fallis certainly wants to avoid, given that he has repeatedly argued that proviso-lies are genuine lies (Fallis 2009, Fallis 2013, Arico and Fallis 2013). Overall, it seems that ACF3 is better interpreted as capturing utterances expressing both partial and outright beliefs– even if this means that it will incorrectly count bets, guesses and other speech acts of the sort as lies.

Stokke’s ACS makes more promising predictions. Recall that for Stokke, *unofficial* common grounds (CGs) are temporary CGs that open up in order to store an information that is used for the purpose of an argument – as opposed to “permanent” *official* CGs. This distinction allows for assertions (and perhapspromises and assurances like (1-2)) to count as lies, and we have already seen that it successfully discards assumptions like (5), as assumptions are only meant to be stored in the *unofficial* CG.

Despite its success in getting rid of assumptions, it seems that ACS is not able to deal with all explicit performatives correctly. On the one hand, it is underdetermined: without a systematic account of how performative utterances contribute to updating the CG, the predictions of ACS in cases like orders or bets (4,6) are unclear. In this sense, absent some refinements, ACS fails to solve the second speech-act theoretic puzzle. On the other hand, even if we supplement ACS with Stokke’s (2014) account of how *some* performative utterances contribute to update the CG, the resulting criterion gives the wrong predictions in many cases.

For instance, according to Stokke (2014) imperatives contribute to the CG by adding a speaker intention to it: in this sense, (4a) adds (4b) to the CG:

(4a) Clean the toilet!

(4b) I want you to clean the toilet

(4) I order you to clean the toilet (4\*) [that you will clean the toilet]

Insofar as we consider (4) to be in some important sense equivalent to (4a) (which is needed to extrapolate a refinement of ACS from Stokke 2014), we have that ACS predicts that (4b) is a lie whenever it is uttered by a speaker that believes (4b) to be false. This is incorrect, however, as this sort of orders can perhaps be misleading, but they cannot be lies.

A way to resist this objection is to interpret ACS more literally: according to a strict version of ACS, the proposition that we identify as the *content* of the performative utterance needs to ‘match’ the proposition added to the CG (it needs to be the same proposition). In (4), there is a mismatch instead: rather than the content of the utterance (4\*), it is (4b) that is added to the CG. Therefore, the strict version of ACS does not count orders like (4) as lie.

This strategy perhaps helps with orders, but prohibiting this sort of mismatches leads to further problems. To see this, consider a denial like (7):

1. I deny that (7\*) I ever saw her kissing your husband

In uttering (7), the speaker is proposing to add the negation of (7\*) to the common ground, so that there is a mismatch between the content of the utterance (7\*) and the proposition that the speaker is proposing to add to the common ground (the negation of (7\*)). Denials like (7) can clearly be lies, but if ACS is interpreted as only capturing ‘matching’ propositions, denials like (7) fall out of its reach. Overall, it seems that every version of ACS encounters serious problems: unrefined, its predictions are too vague; when refined following Stokke (2014), it is too broad; when restricted to only count ‘matching’ propositions, it is too narrow.

What about Carson’s view? ACC is on the right track. In assuring or promising something, a speaker warrants that what he assures or promises is true. And while orders like (4) suggests that (4\*) will probably be the case, it seems clear that in uttering (4) a rational speaker cannot *intend* *to* *warrant* that (4\*) will be true, if warrant is understood as an implicit promise. A similar point applies to bets: while they can suggest that a proposition (e.g. (6\*)) is true, in betting that (6\*) is true a speaker cannot rationally intend to assure their audience that (6\*) is true, because bets are only felicitous in contexts in which the truth of the proposition at stake is uncertain. It seems that, overall, Carson’s account is the only one to make the right predictions when it comes to performative utterances. Nonetheless, even if ACC is on the right track, Carson’s account as a whole (like any other developed so far) still fails to treat all performative utterances correctly, because of the third speech-act theoretic puzzle. As we will see, the third problem with standard account is that not all performative utterances need to be believed to be false in order to count as lies.

### 2.3. Insincerity Conditions

Under which conditions is a performative utterance a lie? According to all definitions, an utterance is a lie only if it meets the insincerity condition (b), *i.e.* if its content (identified in a *non-descriptivist* fashion) is believed to be false. Call this view, that is endorsed by virtually every author in the literature[[42]](#footnote-42), the ‘only-belief’ view of the insincerity condition for lying.

ONLY-BELIEF: the only attitude (or lack thereof) relevant to determine whether an utterance is a lie is belief

Speech act theorists tend to have a broader view of what makes an utterance *insincere*. On their view, believing that the content of an utterance is true is not always sufficient (nor necessary) for that utterance to be sincere. Consider again the case of promises and recommendations:

1. I promise that (2\*) I will clean the toilet today
2. I recommend that (8\*) you bring Damiano to the party

The standard view in speech act theory is that insincerity is not necessarily a matter of belief: other mental states can be relevant to determining whether a speaker is insincere. To consider these two specific cases, scholars tend to agree that a promise is insincere when the speaker does not *intend* to perform the act (e.g. does not intend to (2\*)), and a recommendation is insincere when the speaker does not *desire* that the interlocutor acts on the recommendation (i.e., in (8), if I do not want you to bring Damiano to the party).

Now, even if speech act theorists are right that insincerity is not always a matter of belief, this is not *ipso facto* a problem for standard definitions of lying. Sentences like (8) cannot count as lies: for the purpose of defining lying, we can ignore them, so that it does not matter if they are insincere under different conditions. And even if it seems right that intending to perform the promised act is required for sincerely promising, it might be argued that whenever you *intend* not to perform the promised act you also *believe* that you will not perform it, so that the standard insincerity conditions will capture all insincere promises. This suggests that, for the purpose of defining ‘lying’, the standard insincerity condition might as well do the job.

This reply is on the right track with respect to recommendations, but it is not clear that it solves the problem of promises, as intentions and beliefs do not always converge in the envisaged way. Surely in the majority of cases one’s intention to *p* is accompanied by the belief that one will *p* (and vice versa). However, as stressed by Marušić (2013:293), we can utter promises where intentions and beliefs come apart. We often promise “to exercise more, to get out of debt, to quit smoking, to learn a foreign language, or to spend the rest of our life with our spouse” in situations in which we both intend to stick to the promise and *believe* that we will probably end up violating it.

Intentions and beliefs can also be divergent in the opposite way: it is possible to believe that we will end up doing something despite intending not to do it. To mention cases from the previous list, we can promise *to continue exercising, to avoid smoking again, and not to cheat on our spouse* in situations in which we do not intend to stick to the promise, but believe that we will probably end up respecting it nonetheless. To see how all this relates to lying, consider the following example.

Unfaithful Wife

Baba and Coco are a married couple. Baba is away from home for work, and is planning to go out for a couple of drinks. Since Coco is extremely jealous, during his phone call with Baba he asks her: “Will you be cheating on me tonight?”. Baba replies:

1. Do not worry Coco: I promise that I will not cheat on you tonight

In fact, Baba intends to do her best cheat on Coco tonight, but she is virtually certain that she will end up not doing so, as her terribly awkward manners have always prevented her from seducing any man other than Baba.

Now, in the example Baba has an insincere intention, and her promise is deceptive: it seems intuitive that (9) is a lie. Nonetheless, the *only-belief* view predicts that this is not a lie.

It might be replied that it is not clear that (9) is a lie, because there is *some* component of sincerity in Baba’s utterance. Surely there is *some sense* in which Baba is sincere; the question at issue, however, is whether intuitively our ordinary concept of ‘lying’ applies to this case. It seems to, and there is decisive empirical evidence that this is the case: native English speakers overwhelmingly classify this case and similar others as lies (Marsili 2016, reported in IV.3). Against the predictions of traditional definitions, some utterances can be lies despite their content being believed to be almost certainly true. As I will argue in IV.3, the best way to treat these cases is to allow for propositional attitudes other than beliefs to satisfy the insincerity condition – that is, allow them to determine whether one’s utterance is a lie.

The three puzzles considered so far have shown that traditional definitions of lying fail to handle performative utterances in a satisfactory way, and that every definition needs serious revisions to deal with these cases correctly. First, the statement condition must be amended to allow for a non-descriptivist reading of explicit performatives; second, the third additional condition (whether deceptionist or assertion-based) must be revised to make the right distinctions between performative utterances that can and cannot count as lies (with the exception of Carson’s ACC, that seems to make the right predictions); third, the insincerity condition must be amended to include attitudes other than belief to determine the insincerity of the utterance. In the next section, I develop a speech-act theoretic definition that aims to meet the first two of these requirements. I will deal with the revised insincerity condition in chapter IV, in which I consider issues of insincerity more in detail.

## 3. A Speech-act Theoretic Account of Lying

### 3.1 The Assertion Condition

At least since Frege’s *Begriffsschrift* (1879)*,* there has been growing consensus among philosophers that assertion is a particular kind of *force* – namely, assertoric force. In this context, the term ‘force’ (or ‘illocutionary force’) indicates what a speaker does in virtue of uttering a sentence (as opposed to what he says): for instance, *asking* a question, *issuing* an order, *promising* to do something. Speech act theory is a theory of forces; if defining lying requires defining assertoric force as opposed to other forces, it is quite natural to resort to this theory to develop such an account.

Different positions about assertions can be found in the speech-act theoretic literature. Some authors are ‘fundamentalist’ about assertion (they believe that assertion is the most fundamental, or primitive, of all forces, as in Brandom 1994), others are not; some endorse normative accounts (they believe that assertions are identified by normative constraints, as in Searle 1969 or Williamson 2000), others do not (e.g. Bach & Harnish 1979). Despite these differences, however, there is almost universal consensus that assertions have *word-to-world* direction of fit[[43]](#footnote-43), and several authors (Searle 1969, Searle & Vanderveken 1985, Brandom 1983, 1994, Alston 2000, Green 2000, McFarlane 2011) agree that to assert is to take responsibility for the truth of a proposition by stating that proposition. My proposed definition borrows its three key conditions from this tradition; it can be phrased as follows:

**Definition of Assertion**

A speaker S successfully asserts that p iff:

1. S expresses p
2. S presents p as an actual state of affairs
3. S takes responsibility for p being an actual state of affairs

**Definition of Lying**

S lies iff S successfully and insincerely asserts that p

A few qualifications are in order. The first is that this is an account of intentional assertion, and that all conditions are taken to be satisfied *intentionally* by the speaker. This is a standard procedure in speech act theory (cf. Searle 1969, Alston 2000, but cf. Alston 2000:137-41), and is particularly unproblematic here given that my ultimate explanandum is lying, and virtually every author (cf. A.1) agrees that there can be no such thing as *unintentional* lying[[44]](#footnote-44).

Condition (a) *(S expresses p)* is equivalent to the statement condition encountered in traditional definitions of lying, when such condition is understood in a non-descriptivist fashion. ‘Express’ is a term of art employed by Searle to indicate that the speaker has performed a speech act with content *p*[[45]](#footnote-45)*.*

Condition (b) *(S present p as true)* should be understood as a shorthand for the claim that assertion has a *word-to-world* direction of fit (Searle 1976). That speech acts have different directions of fit, and that assertions (as well as other ‘assertive’ speech acts) have *word-to-world* direction of fit is currently orthodoxy in speech act theory (e.g. Searle 1969, Searle & Vanderveken 1985, Recanati 1987:147-63, Vanderveken 1990, Green 2013), and more generally in philosophy of language and linguistics. Despite the agreement on this view, condition (b) rarely features in definitions – perhaps because its shared by other illocutionary acts. This condition plays nonetheless an important descriptive role, for the very reason that it places assertion within a system of logical relations with other illocutionary acts: this direction of fit is a feature that assertion shares with all the illocutionary acts that belongs to the ‘assertive’ class (acts like reminding, guessing, testifying, assuring, warning, and so forth) but not with other families of speech acts (cf. section 4.2 for more comprehensive discussion).

Condition (c) is somehow self-explanatory, but it needs some elaboration, as it does the lion’s share in the definition. The notion of taking responsibility, also referred to as ‘commitment’ (cf. Searle 1969, Brandom 1983; 1994, MacFarlane 2003a, 2003b, 2005, 2011), is analogous to Carson’s notion of warrant, except for the fact that it is not construed in analogy with promises. The question we need to ask is: what kind of responsibility is exactly in place when one takes responsibility *for the truth of a proposition*?

Like Carson’s notion of warranting, the notion of taking responsibility has to do with the speaker accepting to be reproached if the proposition turns out to be false. An early formulation of this idea is found in Pierce: “an act of assertion […] renders [the speaker] liable to the penalties of the social law (or, at any rate, those of the moral law) in case [the asserted proposition] should not be true, unless he has a definite and sufficient excuse” (CP 2.315). Alston (2000:55) offers a more detailed explanation: in uttering a sentence, a speaker takes responsibility for *p* iff the speaker “knowingly took on the liability to (laid herself open to) blame (censure, reproach, being taken to task, being called to account), in case of *not-p*”[[46]](#footnote-46).

Authors like Brandom and McFarlane correctly point out that being committed to a proposition also involves some eminently *discursive responsibilities.* For instance, according to MacFarlane (2003b), in asserting that *p,* the speaker accepts to:

* 1. be responsible for providing adequate grounds for *p* in response to any appropriate challenge, or
  2. defer this responsibility to another asserter on whose testimony he is relying, or
  3. discharge his commitment by withdrawing the assertion.

To sum up, the notion of commitment (or taking responsibility) can be explained in terms of two components, that for brevity we can call ‘accountability’ (liability to reproach) and ‘discursive responsibilities’ (i-iii)*.* One is committed to *p* when these responsibilities arise *in virtue of* one’s utterance that *p.* In terms of defining lying, this notion purports to do the same job of the notion of warrant introduced by Carson, but it avoids the problematic implications of defining these responsibilities in terms of an analogy with promises. Before considering whether this definition is successful in addressing the puzzles introduced in section 2, I want to explain how this account relates to a general theory of illocutionary forces, and which kinds of distinctions it introduces between assertions and other illocutionary acts.

### 3.2 Illocutionary Entailment

I have already suggested that it is possible to perform more than one illocutionary act at the time. For instance, it seems that when I promise that I will do something, I am also thereby asserting that I will do that thing. Speech act theory provides excellent tools to analyse this sort of relationships between illocutionary acts – relationships that are here captured under the label of ‘illocutionary entailment’.

According to Searle & Vanderveken, the standard way for a speaker to successfully perform an illocutionary act F(p)[[47]](#footnote-47) is to utter a sentence that expresses literally that force and content: this is called a *literal performance* of that act. In this sense, (1) is a literal performance of a promise.

(1) I promise that (1\*) [I will feed the brontosaurus]

Literal performance is to be contrasted with two kinds of *non-literal* performance. First, there are cases of non-literal performance due to entailment: here, an illocutionary act *entails* the performance of another (1985:129-30). The performance of an illocutionary act F1(p) entails the performance of another illocutionary act F2(p) iff in the context of the utterance it is not possible for S to perform F1(p) without performing F2(p) – so that if S performs F1(p), S also performs F2(p). In this sense, my explicit promise (1) illocutionarily entails the assertion that I will feed the brontosaurs, since I cannot promise that I will feed the brontosaurus without thereby also asserting that I will feed the brontosaurus.

Second, there are non-literal performances of F2(p) that the speaker performs by means of performing another illocutionary act F1(p) that does not entail F2(p), but in the performance of which he makes salient his *reflexive intention* (cf. Grice 1989, Bach & Harnish 1979, and II.5.2.3) to make the hearer understand that he intends to perform F2(p) (either instead of, or in addition to, F1(p)). These cases involve standard implicatures, such as *conversational implicatures* or *scalar implicatures* –as the one, for instance, from (2) to (2a):

(2) I ate three bananas

(2a) I did not eat more than three bananas

According to this distinction, and consistently to the predictions of my account, *entailing* a proposition can qualify as asserting, whereas implicating a proposition cannot. My account requires (condition (a)) that the speaker express the proposition that he asserts. For this reason, when the other conditions (b)-(c) obtain, any entailed act can qualify as an assertion, whereas an implicated proposition cannot. To go back to our example, in uttering (1) I express the proposition (1\*) that I will feed the brontosaur, thus satisfying condition (a), whereas in uttering (2) I do not express the proposition (2a) that I did not eat more than three bananas, thus failing to satisfy condition (a).

The prediction that implicatures will not count as lies is desirable, as it is consistent with the traditional view that lying requires saying something, and that implicating a proposition that one has not literally expressed amounts to misleading rather than lying. In the forthcoming section, I will show that the notion of entailment gives us a good criterion to apply the definition of lying to illocutionary acts other than assertion, to determine which ones (under normal conditions) can entail an assertion (e.g. promising, assuring, reminding or swearing that *p)* and which cannot (e.g. ordering, congratulating or betting).

## 4. Testing the definition

How does the definition deal with the problematic cases introduced in the first half of the chapter? I shall show that neither the counterexamples to assertion-based definitions (section 1) nor the speech act theoretic puzzles (section 2) pose a threat to the definition. Let us consider them in turn.

### 4.1 Proviso Lies

With respect to counterexamples to assertion-based definitions, only those targeting Carson’s proposal are relevant: my definition of assertion differs so significantly from Stokke’s and Fallis’ that it would be superfluous to show that counterexamples to these accounts do not apply to mine[[48]](#footnote-48).

Carson’s proposal incorrectly rules out *proviso lies*: believed-false statements followed by provisos that remove the warrant that the proposition is true; for instance: “*p*, but I might be wrong”, or “*p,* but do not take my word for it”. I have argued that Carson’s cannot claim that provisos mitigate rather than remove warrant, because he understands warrant as an implicit promise. You cannot mitigate an implicit promise, because promises do not come in degrees: either one promises that *p*, or one does not.

This difficulty does not arise for the notion of commitment used in my definition. The possibility of strengthening or diminishing the speaker’s degree of commitment to a proposition is widely acknowledged in the speech act theoretic literature (Searle 1976:5, Holmes 1984, Searle and Vanderveken 1985:98-9, Coates 1987: 112, Bazzanella, Caffi & Sbisà 1991, Sbisà 2001:1805-06), and actually plays an important role in explaining the relations of entailment between different illocutionary acts (Searle & Vanderveken 1985:130-1). For instance, authors agree that in choosing to use the performative ‘swear’ in (4a) (instead of plainly asserting (4)) the speaker strengthens his commitment to the proposition (4\*), whereas in choosing the performative ‘guess’ in (4b) he removes such commitment.

1. Jimmy was drunk yesterday night

(4a) I swear that (4\*) [Jimmy was drunk yesterday night]

(4b) I guess that (4\*) [Jimmy was drunk yesterday night]

All three speech acts belong to the class of assertives, and present the proposition (4\*) as true. Since (4a) involves a stronger commitment than assertion, its utterance illocutionarily entails the performance of an assertion – it cannot be performed without thereby asserting that (4\*) is true. By contrast, since (4b) involves no commitment to the truth of (4\*) (the speaker cannot be reproached if Jimmy was not drunk), its utterance entails no assertion.

This shows that commitment can be *reinforced* or *removed*, but not yet that it can be *mitigated* while still asserting, which is what we need in order to prove that my definition can acknowledge that proviso lies are genuine lies. Cases of this sort, however, are not uncommon, and typically emerge from the use of some modifiers, such as *evidentials* or *epistemic modals.* For instance:

(4c) Apparently (4\*) Jimmy was drunk yesterday night

In uttering (4c), I accept some responsibility for the truth of what I have said. For instance, it would not be inappropriate for someone to reply “How do you know?” to (4c), whereas it would be inappropriate to offer this reply to my guessing that Jimmy was drunk in (4b). The same story seems to apply to the proviso-lie encountered in section 1.1:

1. Mel broke your trophy. But I was kinda drunk, and there were lots of people in there, so don’t take my word for it.

Also in this case, the speaker is signalling that he is not willing to accept full responsibility for the proposition being true. Like the mitigated assertion (4c), and unlike the case of betting (4b), it is appropriate to inquire about the ground for the speaker’s assertion, and to reproach him if the assertion turns out to be false[[49]](#footnote-49): both ‘accountability’ and ‘discursive responsibility’ are satisfied. This shows that proviso lies and mitigated lies still count as assertions (and therefore as lies) according to my definition, and that this verdict is consistent with verdicts given in similar cases like (4c).

### 4.2 The Speech-act Theoretic Puzzles

What about the speech-act theoretic puzzles? The first puzzle is obviously not a problem, as it demands a non-descriptivist semantics for performative utterances: my account naturally incorporates such semantics – unlike others account, that needed to be adjusted *ex-post*, in light of their failure to deal with performative utterances.

The second puzzle is more complicated to solve, as there are countless illocutionary acts, and it is difficult to prove that any account deals correctly with each and every one of them. Nonetheless, I will try to demonstrate that for each *family* of speech acts, my account gives the correct predictions as to whether its members can or cannot count as lies. The following scheme represents a standard classification of speech act in five families, based on their direction of fit[[50]](#footnote-50):

|  |  |  |  |
| --- | --- | --- | --- |
| Class | DOF | Illocutionary point | Examples |
| Assertives | ↓ | commits S to the truth of p | Assert, inform, suppose |
| Commissives | ↑ | commits S to do p | Promise, accept, vow |
| Directives | ↑ | S attempts to make H do p | Order, request, advise |
| Expressives | ∅ | S expresses a feeling | Congratulate, thank |
| Declaratives | ↕ | S makes it the case that *p* solely in virtue of the successful utterance of F(p) | Declare, christen, fire, resign, excommunicate |

Table 1: a taxonomy of illocutionary act in terms of direction of fit, as presented by Searle (1976)

Let us consider these categories in turn.

• Assertive illocutionary acts can count as lies whenever the speaker is committed (at least to a sufficient extent) to the truth of the proposition, *i.e.* whenever the speaker takes responsibility for the truth of the proposition. This means that almost all assertives (reminding, testifying, reporting, assuring, etc.) can count as lies, with a few exceptions, like *conjecturing, supposing* or *guessing,* that fail to meet the ‘accountability’ component of assertoric commitment (for more extensive discussion of this point, cf. Marsili 2015:124)[[51]](#footnote-51).

• Directives have an opposite direction of fit, so that they cannot meet condition (b) of my definition of assertion (the speaker presents *p* as true) and they cannot count as lies. This is a welcomed prediction, as there seem to be no conditions under which a command, an advice or a request can count as lying.

• Commissives also have a different direction of fit – *qua* commissives, they cannot count as lies. However, unlike with directives, in virtually every case the performance of a commissive illocutionary entails the performance of an assertion – as in the well-discussed case of promises. Therefore, these speech acts can count as lies. An important difference from assertives is that commissives can be insincere also when the speaker lacks an intention, not only when he lacks a belief (see IV.2). Furthermore, since commissives always require the speaker to undertake responsibility for the truth of a future state of affairs, there are no commissive that involve null responsibility (Searle and Vanderveken 1985:100), as it happens with some assertives, like guessing. This means that unlike assertives, all insincere commissives count as lies.

• Expressives are a bit trickier. Traditionally, they are taken to have no direction of fit, as they do not explicitly express a full proposition. For instance, in saying “Thanks!” or “Congratulations!” a speaker does not explicitly express a proposition that can be true or false, or that he can believe to be true or false. As such, expressives cannot meet condition (1) or (2) of my proposed definition of assertion, and therefore cannot count as lies.

Nonetheless, there seem to be cases that do not fit this description. Quite often, expressives are performed by means of assertions. For instance, the assertion (5) is a compliment, and (6) is a congratulation:

1. The cake was delicious
2. Congratulations, your paper was great!

Utterances like (5) and (6) can clearly count as lies. But this is not a problem for my proposal. These cases are simply the converse of speech acts entailing assertions (as in the case of asserting by promising): they are assertions that entail another illocutionary act, namely an expressive. In claiming that expressives cannot count as lies, I am not claiming that assertions cease to be assertions if they entail an expressive. My claim is simply that a *mere* expressivecannot be a lie; whenever an expressive is performed by means of an assertion, it can obviously be a lie in virtue of its being, first and foremost, an assertion. To be sure, this is quite different from the case of commissives entailing assertions: commissives entail assertions by default, so that they can always count as lies, whereas expressive are only occasionally performed via assertions, and only in such cases they can be lies.

• Declaratives are the trickiest of all classes. They share the direction of fit of assertives, but also that of directives: their point is to change the world *in virtue* of describing the world as being in a certain way. In principle, this does not prevent them from counting as assertions, as in some sense they do present a proposition as being true[[52]](#footnote-52). Our concern here, however, is which illocutionary acts can count as lies. Declarations could be easily disqualified even if we count them as assertions, as they possess the special feature of being true in virtue of their being performed successfully. This makes it very difficult for a speaker to perform a declaration insincerely, because it is difficult for a speaker to believe that a sentence that is true in virtue of its being uttered is false. We may conclude that they can never count as lies, because they can never meet the relevant insincerity condition.

Such a conclusion, however, would be too hasty. Even if it is difficult for a speaker to believe his declaration to be false, it is not impossible. There are some rare cases in which a speaker can believe his declaration to be false; arguably, only two types of declarations of this sort. Before we consider them, it is worth noting that since these cases are anything but ordinary, different people will have different intuitions in these cases, and weak intuitions rather than strong ones. If this observation is correct, we should not consider such intuitions to be decisive to resolve matters in favour of one definition of lying against the another. Still, it will be interesting to see what my account has to say about these trickier cases, and to consider to which extent its predictions are able to acknowledge that these cases are less straightforward than others.

First, there are declarations that are covertly intended to fail: here the speaker secretly believes that his declaration will be unsuccessful, typically because he lacks the necessary authority. For instance, imagine the case of a couple who is tricked into a staged marriage. The priest is not really a member of the clergy, and he knows that the ceremony will be invalid, both in the eyes of the Church and in the eyes of the law. As the fake marriage unfolds, after a few ceremonial sentences, the fake priest says:

(7) I now declare you husband and wife

Has the priest lied? My intuition is that (7) is not a lie, even though it possesses several features that lies typically have. My definition supports this intuition: it does not classify (7) as a lie, but at the same time it is acknowledges its borderline nature. Let us see how. On the one hand, there is a sense in which the priest meets my assertion-condition (b) of presenting the expressed proposition (that the couple is married) as being true: he presents it as true *in virtue of* his saying so. On the other hand, I made it explicit that the requirement (b) that an assertion presents a proposition as true is a shorthand for requiring a word-to-world direction of fit. I follow speech act theorists in meaning that “the propositional content of the illocution [must] fit an *independently* existing state of affairs” (Searle & Vanderveken 1985:53). Understood in this strict sense, the condition is not met by the declaration, because the declaration *alters* the relevant state affairs by representing that state of affairs as being so altered.

A similar mixed verdict arises about condition (c), taking responsibility for the proposition being true. On the one hand, there is clearly a sense in which ‘accountability’ is satisfied: the couple can reproach him for stating (7), were they to discover that they are not married. On the other hand, it is not clear that ‘discursive responsibility’ is satisfied. It would be inappropriate for someone to ask to the priest “How do you know that they are married?”, and the priest does not need to provide “adequate grounds for [his declaration] in response to any appropriate challenge, or defer this responsibility to another asserter on whose testimony one is relying”. What can be challenged is rather the priests’ authority to perform the marriage, and the priest is only supposed to defend his *authority* to utter the declaration in that context. Even if the utterance of a declaration involves undertaking some responsibility, it is clearly a different kind of responsibility, that elicits different challenges and demands different defences.

The bottom line is that while my definition acknowledges that (7) has several features in common with an assertion, it does not count it as one unless we loosen both condition (b) and (c) – in line with my intuitions. If one has the opposite intuitions, a broader understanding of ‘discursive responsibility’ and of ‘presenting *p* as true’ can be adopted; absent these refinements, however, the definition does not count (7) as a lie.

A second kind of declaration that can in some sense be believed to be false involves Austin’s (1962:152) category of verdictives. Verdictives are declarations that institutionally sanction an independently existing state of affairs, like a referee’s sanction of an offside, or a judge’s verdict regarding a given crime. Since such verdicts can be at odds with reality, there is a sense in which they can be believed to be false. To see if they can also count as lies, let us consider an example: the case of a judge that has been bribed to support a jury’s verdict. The jury has established that the defendant is guilty; the judge knows that the defendant is innocent, has the power to challenge the jury’s decision, but nonetheless decides to support it to cash the bribe. Can his utterance (8) be considered a lie?

(8) I declare the defendant guilty

In this case, too, intuitions are not straightforward. Like in the previous example, my definition predicts that this is not a lie, while acknowledging that it is a borderline case. On the one hand, it seems that the judge says something he believes to be false: he believes that the defendant is innocent, rather than guilty. On the other hand, the judge clearly believes (8) to be true, because the defendant is now guilty in the eyes of a law, in virtue of his saying so.

This mixed analysis results from two different meanings of ‘guilty’ being considered: guilty of committing the crime, and guilty in the eyes of the law. Now, it seems that the relevant meaning of ‘guilty’ is the latter: the judge is declaring that the defendant is ‘guilty’ in the eyes of the law. Clearly, the judge should not make this declaration if he knows that the defendant is not also guilty of committing the crime, so that in uttering his declaration the judge is also implying or presupposing that the defendant is also guilty in the former sense. What is at stake here, however, is whether the judge has lied in uttering (8). And in implying or presupposing that the defendant has committed the crime, the judge is not saying that the defendant has committed the crime. Once again, we are dealing with a case of misleading rather than lying. Since the judge does not believe that what he has said is false, the verdict of my definition (and, in this case, of all the others) is that this utterance is deceptive, but not a lie.

## 5. Conclusions

This chapter has reviewed *non-deceptionist* account of lying, more specifically *assertion-based definitions* of lying. I have argued that all the major accounts on the market are subject to counterexamples specific to those accounts. On top of this, I have shown that both *deceptionist* and *non-deceptionist* accounts are unable to solve my three speech-act theoretic puzzles about lying. I then proceeded to introduce my own definition, that can be summarised as follows:

**Speech-act theoretic definition of lying**

In uttering U, S lies to A about p iff:

1. S thereby asserts that p, i.e.:
2. S expresses p
3. S presents p as an actual state of affairs
4. S takes responsibility for p being an actual state of affairs
5. S is insincere about p

This definition avoids the counterexamples that affect alternative assertion-based definition and, unlike other definitions, is able to solve the second speech-act theoretic puzzle –distinguishing illocutionary acts that can be used to lie from illocutionary acts that cannot. Broadly, it predicts that assertives and commissives can count as lies, whereas directives and expressives cannot. Declaratives also do not count as lies, but their case is slightly different: in the rare cases in which they are performed insincerely, they share some important features of assertions, even though they fail to qualify as such.

At the current stage, my proposal is only able to deal with two of the three speech-act theoretic puzzles. I have not yet clarified how it can solve the third one, overcoming the mistakes of the *only-belief* view – an enterprise that I undertake in next chapter, in which I discuss more in detail what it means for an illocutionary act other than assertion to be insincere.

# IV. Insincerity

Sincerity is often valued as an important virtue, and insincerity criticised as a vice. We generally trust other people to be sincere, and their testimony is a fundamental source of information without which we could hardly get on with our ordinary lives. For these reasons, insincerity has elicited the interest of philosophers working not only on language, but also on ethics and epistemology. Any epistemological or ethical discussion of insincerity, however, presupposes settlement of one fundamental question: what is it to be insincere?

This chapter is devoted to answering this question. More specifically, it is concerned with two related aims. Its primary aim is to provide an analysis of insincerity as a component of lying. In other words, the primary goal of this chapter is to refine the insincerity condition in the definition of lying, specifically in the light of some counterexamples to which the traditional definition falls victim, including the third speech-act theoretic puzzle introduced in the previous chapter. But even if it is the interest in lying that drives this enquiry, my analysis of insincerity as a necessary condition for lying arguably retains value also as an analysis of insincerity in general. My related, secondary aim is thus to provide a characterisation of insincerity that (at least to some extent) applies also outside the debate on defining lying.

This chapter is divided in four long[[53]](#footnote-53) sections. The first one simply lays out the problem of defining insincerity, and clarifies more accurately which notion of insincerity I am after. In section 2, I attack the *only-belief view*, namely the idea that beliefs are the only attitude that is relevant to determining whether an utterance is a lie. As I have argued in the previous chapter, this account of the insincerity condition for lying is inaccurate: for instance, a promise can be a lie when the speaker *does not intend* to perform the promised action. I consequently expand the insincerity condition to attitudes other than belief, thereby solving the last of my three speech-act theoretic puzzles. In the section 3 I test my revised condition empirically, showing that ordinary speakers share the intuition that also intentions can determine whether an utterance is a lie.

Section 4 criticises the ‘dichotomic view’ of insincerity, namely the idea that either a statement is believed to be true, or believed to be false. It introduces lies that fall outside this dichotomy, namely *fuzzy-lies* and *graded-belief lies*, and develops a refined insincerity condition that treat these graded lies correctly: broadly, a speaker is insincere if he believes his statement to be more likely to be false than true. The final, fifth section brings together all my findings into a general account of the insincerity conditions for lying.

## 1. Insincerity: a preliminary account

In ordinary language, the terms ‘sincere’ and ‘insincere’ are used in different contexts with different meanings. Before initiating a more thorough discussion of insincerity, I would like to clear up some ambiguities about these different meanings, to explain exactly which notion of insincerity this chapter aims at analysing.

First, in ordinary language insincerity need not refer to linguistic utterances. We can say that a *smile*, or even *person* (as opposed to an utterance) is insincere. While these uses of the term are certainly appropriate, they are not the object of our interest here. This thesis is concerned with insincere utterances and more specifically assertions, and consequently only with linguistic insincerity.

Second, even when we limit our analysis to linguistic insincerity, it seems that this term can be used at least in two ways[[54]](#footnote-54). In a broad sense, calling an utterance insincere is describing it as *deceptive*, or *aimed at deceiving.* For instance, Bernard Williams (2002: 74) defines insincere assertions as those that “have the aim of misinforming the hearer”. Under this conception, an insincere assertion is one intended to deceive: not only lies, but also *misleading* but literally true statements, *omissions*, and any sort of deceptive statement.

This is not the conception of insincerity I am concerned with, for at least two reasons. First, I have already stated that I am after a notion of insincerity that, paired with a notion of assertion, will provide us with all the notions required for defining of lying. Understanding ‘insincere’ as synonymous to ‘deceptive’ or ‘intended to deceive’ would not help in this enterprise: it would not allow for a distinction between lying and merely misleading (cf. A.1.2), and it would conflate the insincerity condition for lying with the intention to deceive condition. Second, on this conception the analysis of “S was insincere (in saying *p*)” would be equivalent with the analysis “S was deceptive (in saying *p*)”. Between a notion that overlaps with another and one that does not, the latter is clearly more appealing, as it enriches our conceptual toolbox in a way that the former does not.

The sense of insincerity in which I am concerned is then a different one. Under this conception, insincerity indicates a discrepancy between the psychological state of the speaker (*e.g.* believing, intending, desiring) and the psychological state expressed by his speech act (*e.g.* asserting, promising, requiring). Defining ‘insincerity’ amounts to defining the nature of this discrepancy, which will be the subject of this chapter.

Finally, insincerity is a complex phenomenon, and there are several philosophical problems that concern this notion, more than this chapter could potentially discuss and analyse. In particular, here I will focus on two problems that have been rarely, if ever, discussed in the literature: how a definition of lying can deal with attitudes other than belief, and with graded insincerity. To focus on these relatively new problems, I will leave aside some classic ones. One in particular will be not be discussed in detail: that of differentiating between insincerity on the one hand, and misspeaking and self-deception on the other. I have briefly addressed this problem in A.1.2, where I argued this distinction can be made explicit just by specifying that the speaker has to satisfy (any version of) the insincerity condition *deliberately.* In order to focus on other issues, I will leave further philosophical problems concerning misspeaking and self-deception aside[[55]](#footnote-55)*.*

## 2. Beyond belief: insincerity and illocutionary acts

In the previous chapter**,** I have mentioned that there seems to be universal agreement in the literature on lying that whether an utterance is a lie, and therefore insincere, is only a matter of what the speaker believes. I called this view, that is endorsed by virtually every author in the literature, the ‘only-belief’ view of the insincerity condition for lying.

ONLY-BELIEF: the only attitude (or lack thereof) relevant to determine whether an utterance is a lie is belief

In III.2.3, I have argued that this view is wrong. I will now readdress such criticism more in detail, and show how the insincerity condition for lying can be expanded to other propositional attitudes (e.g. intentions, desires) in order to address this objection. I will develop an alternative account of what it is for a speech act to be sincere or insincere, and then put it to work against counterexamples based on insincere promises. In the next section, I will show that this account better reflects ordinary speaker’s intuitions about what counts as a lie.

### 2.1 The third speech act theoretic puzzle

Let us start by briefly recapitulating the speech act theoretic puzzle introduced in III.2.3. The example is meant to show that there can be cases of lying in which the speaker believes that the propositional content of the utterance (identified in a non-descriptivist fashion) is true, and in which the speaker’s intentions, rather than his beliefs, are relevant to determine whether his utterance is a lie.

Unfaithful Wife

Baba and Coco are a married couple. Baba is away from the city for work, and is planning to go out this night. Since Coco is extremely jealous, he asks her: “Will you be cheating on me tonight?”. Baba replies:

* + 1. Do not worry Coco: I promise that I will not cheat on you tonight

In fact, Baba intends to do her best cheat on Coco at the party, but she is virtually certain that she will end up not doing so, as her terribly awkward manners have always prevented her from seducing any man other than Baba.

In the example, Baba has an insincere intention, and her promise is deceptive: it seems intuitive that (1) is a lie. Nonetheless, the *only-belief* view incorrectly predicts that this is not a lie, because Baba is almost certain that she will not cheat on Coco. What is missing here is an intention (the intention to try to stick to the promise) rather than a belief. Against the predictions of traditional definitions, some utterances can be lies despite their content being believed to be true.

To avoid this counterexample, a definition of lying should allow for propositional attitudes other than belief to determine whether one’s utterance is a lie. But how are we to extend the insincerity condition to other attitudes? Speech act theory offers a promising theoretical framework for this purpose: it is a standard view in speech act theory that insincerity can depend on a variety of attitudes, including *beliefs, intentions* and *desires*.

### 2.2 Expressing attitudes and (in)sincerity

Broadly put, speech act theorists take a speech act to be insincere whenever there is a mismatch between *the attitude expressed by the utterance* and the *attitude possessed by the speaker* (Falkenberg 1988:93). Taken out of context, this definition is not very informative; in what follows, I will provide some theoretical background to flesh it out in a meaningful way.

It is a standard view in speech act theory that each illocutionary act expresses a distinctive propositional attitude (Searle 1969, Bach & Harnish 1979). The distinct attitude expressed by a given illocutionary act is part of what identifies it as opposed to others, and it is generally taken to define the point or purpose of the actions that we perform in uttering it. In this sense, we say that an assertion expresses a *belief*, that a promise expresses an *intention,* that asking someone for something expresses a *desire*. Philosophers and linguists have presented different taxonomies of illocutionary acts based, amongst other things, on the different psychological attitudes expressed by different (kinds of) illocutionary acts. Most authors would agree that the following characterisation is broadly[[56]](#footnote-56) correct:

**Attitudes expressed by specific illocutionary acts**

B-EX: If S asserts that p, S expresses (EX) the belief (B) that p

D-EX: If S asks for p, S expresses a desire (D) for p

I-EX: If S promises that p, S expresses the intention(I) to do p

There are several ways to flesh out what is meant by ‘expressing’ a psychological state. I will assume that *expressing* an attitude does not entail having that attitude, so that you can insincerely express an attitude that you do not have (but cf. Davis 2003: 25, Green 2007:70-83). Following a hint by Davidson (1985:88, cf. Marušić 2012:13, Fallis 2013), one could say that for a speaker to express a psychological state is for the speaker to *represent himself* as being in that psychological state.

On an orthodox speech-act theoretic account of sincerity[[57]](#footnote-57), the sincerity condition for uttering a given illocutionary act is that the speaker has the psychological attitude expressed by that act:

**Sincere illocutionary acts**

SIN: The performance of an illocutionary act F(p) that expresses the psychological state Ψ(p) is sincere IFF in uttering F(p), S is in Ψ(p)

From the orthodox account of sincerity, a simple account of insincerity can be derived: a speaker is insincere whenever he is not in the psychological state that is expressed by the illocutionary act performed:

**Insincere illocutionary acts**

INS: The performance of an illocutionary act F(p) that expresses the psychological state Ψ(p) is insincere IFF in uttering F(p), S is not in Ψ(p)[[58]](#footnote-58)

This gives us a provisional, simple formulation of the insincerity conditions for the three illocutionary acts we are considering as examples:

**Insincerity conditions for specific illocutionary acts**

BIC: S asserts that p insincerely only if S does not believe that p

DIC: S asks for p insincerely only if S does not desire p

IIC: S promises that p insincerely only if S does not intend to p

Importantly, an utterance can be insincere without being a lie. This is because an illocutionary act can be insincere without satisfying the assertion condition for lying. For instance, in asking you to do something that I do not desire, I may be insincere but I am not thereby lying. It should be thus kept in mind that this is a general account of insincerity, and that without a definition of assertion it does not alone provide us with sufficient conditions for defining lying.

### 2.3 Insincerity conditions for promising

Let us go back to the speech-act theoretic puzzle. The problem introduced by the Unfaithful Wife example is that there seem to be cases in which the speaker lies simply by lacking the intention to perform the promised action. In order to solve this puzzle, we need an alternative account of the insincerity conditions for lying. Since the proposed counterexample is about promising, to simplify this task I will start by developing an account for lying by promising. I will then show that this account generalises to every other illocutionary act that can be used to assert.

At this stage, on the table there are only two competing ways of accounting for the insincerity conditions for lying by promising. The *first* is the approach traditionally used to define lying: the *only-belief* account. This approach applies indifferently the same insincerity conditions to every illocutionary act, and counts an utterance as insincere whenever its content is believed to be false.

The *second* is the speech act theoretic account that I just introduced. According to this account, (i) a promise expresses an intention to *Φ*, and consequently (ii) the insincerity condition for promising is intending not to *Φ*. Assumption (i) can be traced back to Hume’s view (*THN:* 517-19) that a promise always *expresses* (and communicates) an intention to perform the promised act[[59]](#footnote-59). Assumption (ii) is found in foundational works of speech act theory, like those of Austin (1962/1975:50, 135-6) and Searle (1965:243, 1969:60-2). I will refer to this account as the *only-intention* account. On this view, a promise is insincere only if, *at the time of the utterance*, the speaker does not intend to perform.

A *third* account of the insincerity conditions for promising can be derived by combining the previous two, in the light of the relation of entailment between asserting and promising that I outlined in III.3.2. As a reminder, the performance of an illocutionary act F1(p) entails the performance of another illocutionary act F2(p) iff in the context of the utterance it is not possible for S to perform F1(p) without performing F2(p) – so that if S performs F1(p), S also performs F2(p). This relation of illocutionary entailment occurs between assertions and promises: one cannot promise to *Φ* without also performing an assertion that one will *Φ*, so that every time one promises to *Φ* one also asserts that one will *Φ*. To recycle my previous example, my explicit promise (2) illocutionarily entails the assertion that (2\*), since I simply cannot promise that I will feed the brontosaurus without thereby also asserting that I will feed the brontosaurus:

* + 1. I promise that (2\*) [I will feed the brontosaurus]

How does this affect the insincerity conditions for promising? My conjecture is that if an assertion is always performed in addition to a promise, for the promise to be sincere the sincerity conditions for asserting need to be satisfied too.In other words, performing a sincere promise that *Φ* requires one both to intend *Φ* and to believe that one will *Φ*. This yields a novel account of the insincerity conditions for promising:

**Entailed-Insincerity condition:**

A promise is insincere if speaker intends not to Φ, or if the speaker believes that he will not Φ, or both

More specifically, a promise to Φ is insincere *qua promise* if the speaker intends not to *Φ*, and insincere *qua assertion* if the speaker believes that he will not *Φ*. This view (be it correct or not) can clearly be generalised: whenever there is illocutionary entailment, and two illocutionary acts are performed, the sincerity conditions of both acts apply. To sum up, the three candidate insincerity conditions for lying by promising are:

**Candidate insincerity conditions for promising:**

(BIC): Belief insincerity condition: S believes that S will not Φ

(IIC): Intention insincerity condition: S intends not to Φ

(EIC): Entailed insincerity condition: BIC ⋁ IIC

Which account is preferable? The Moorean test for insincerity provides some linguistic data that *prima facie* favours the *entailed-insincerity* view over the other two. It is well known that assertions followed by the negation of their sincerity condition give rise to Moorean absurdities (Moore 1993:210): asserting “p and I don’t believe it” is incoherent in some distinctive way. One way of explaining this incoherence is that in uttering these sentences, the speaker performs a speech act and then blatantly violates one condition for its felicitous (in this case, sincere) performance, eventually failing to assert that *p* (Vanderveken 1980, Searle and Vanderveken 1985: 150-52). If both BIC and IIC are insincerity conditions for promising, it seems that they should both give rise to the same kind of unsuccessful incoherence (cf. Marušić 2012:14). As a matter of fact, sentences like (2¬B) and (2¬I) display this kind of absurdity:

(2¬B) I promise that I will pick you up at 6, but I don’t believe I will pick you up at 6 #

(2¬I) I promise that I will pick you up at 6, but I don’t intend to pick you up at 6 #

In both cases, the utterance strikes one as incoherent, and in both cases, it is difficult to imagine that the speaker will be taken to have promised to pick his interlocutor up at 6. This linguistic data supports EIC, and cannot be easily explained by BIC or IIC taken separately.

More importantly, only the EIC seems to make the right predictions in cases in belief and intention come apart, as in Unfaithful Wife. In Unfaithful Wife, Baba has a sincere belief that she will end up sticking to her promise (1), but an insincere intention to do whatever is in her power to break it.

* + - 1. Do not worry Coco: I promise that I will not cheat on you tonight

While *the only-belief condition* BIC makes the incorrect prediction in this case, the *only-intention condition* IIC and *entailed insincerity condition* EIC correctly count (1) as a lie. The latter two conditions are thus preferable. But unlike IIC, EIC makes the correct predictions also when belief and intention diverge in the opposite way, *i.e.*  when the speaker intends to do his best to stick to his promise, but believes he will end up violating it nonetheless. To see this, let us consider another example:

Unreliable Mechanic

Baba has broken her car, but she needs it to visit her family next week. For this reason, Baba has called Coco the mechanic to repair it. Coco the mechanic checks the car and tells Baba:

* + 1. Do not worry Baba: I promise that I will repair your car by next week

Coco intends to repair the car and he will attempt to do it, but he is almost certain that he won’t manage to repair it in the end, because the damage is too serious.

In this example, Coco the mechanic promises to repair the car even if he knows that he will almost certainly fail to repair it: intuitively, (3) is a lie. In this case, the *only-intention* condition IICis not able to track this intuition; whereas both the *entailed-insincerity* conditionEIC and the *only-belief* condition BIC, correctly predicts that (3) is a lie.

Cases like Unfaithful Wife and Unreliable Mechanic support the conjecture that EIC is a better account of the insincerity conditions than BIC and IIC taken separately. Arguably, these cases are not straightforward, or prototypical, cases of lying. But this is also a prediction of the *entailed-insincerity* account: (3) is insincere *qua assertion*, but not insincere *qua promise.* Coco the mechanic intends to fulfil the promise, following one sincerity condition, but he believes he will almost surely fail, violating the other. By contrast, (1) is insincere *qua promise*, but not *qua assertion.* Baba intends to do her best to break the promise, violating one sincerity condition, but she believes that despite her efforts she will end up violating it, thereby following the other.

All in all, it seems that there are solid reasons to prefer the *entailed-insincerity* account of promising. So far, only Marušić (2013) has defended a similar view (Austin 1961:239 merely hints at this idea). However, Marušić also suggests that a promise that violates BIC but not IIC might be better described as *irrational* rather than *insincere*. This is true if one endorses a ‘cognitivist view’, contending that it is irrational to intend to do something you believe that you will not do, so that only utterances satisfying both BIC and IIC (or neither of them) are rational, while utterances like (3) are *irrational* rather than *insincere*.

Marušić’s observation points out a possible problem for the proposed account: if a rational intention to *Φ* requires believing that one will *Φ*, then there is no need to require both IIC and BIC, as the satisfaction of the first entails the satisfaction of the second in every case in which the speaker is rational. I will not discuss this objection here; however, in Marsili (2016) I have argued that (as long as you take it as a live possibility that you will *Φ ­­– i.e.* as long as you are not certain that you will not *Φ*) you can rationally intend to *Φ* and believe that you will very likely not *Φ* (or vice versa). On this *weak cognitivist* view, Marušić’s observation is not a worry: utterances like (3) are *insincere* rather than *irrational*, and EIC is preferable to the other accounts exactly because it successfully captures these peculiar forms of insincerity. In IV.3, I report empirical evidence showing that native English speakers judge that promises with contrasting intentions and beliefs are mendacious rather than irrational, and that they do not find the contrast between intention and belief involved in these cases to be problematic.

### 2.4 A general account of the insincerity conditions for lying

If my arguments are sound, it should be established at this point that EIC offers the best characterisation of the insincerity conditions for lying by promising. Pairing the EIC with the definition of lying developed in the previous chapter, we can obtain the following definition for lying by promising that *p:*

**Definition of lying by promising**

*In successfully uttering a promise with content p, S lies to A about p iff:*

1. *S thereby asserts that p*
2. *Either S believes that not p, or S does not intend to p, or both*

Since promises by default entail assertions, condition (1) obtains by default: the informative bit is condition (2), that specifies under which conditions a promise is a lie. This definition can then be extended to other illocutionary acts that entail an assertion. We know from the general account of insincerity developed in the previous section that an illocutionary act F(p) that expresses the psychological state Ψ(p) is insincere IFF in uttering F(p), S is not in Ψ(p). To generalise the definition of lying by promising to illocutionary acts other than assertion, we simply need to require that, when an illocutionary act other than assertion is preformed, either this insincerity condition is satisfied, or the insincerity condition for assertion, or both. In other words:

**Definition of lying by performing an illocutionary act**

In successfully uttering an illocutionary act with content p that expresses an attitude Ψ(p), S lies to A about p iff:

1. S thereby asserts that p, i.e.:
2. S expresses p
3. S presents p as an actual state of affairs
4. S takes responsibility for p being an actual state of affairs
5. Either S believes that not p, or S is not in Ψ(p), or both

This gives us a more accurate account of the conditions under which an illocutionary act other than assertion can count as lying. This definition is more accurate than the one developed in chapter III, as it generalises the insincerity conditions to attitudes other than beliefs. But does this definition really reflect our ordinary intuitions about lying? In what follows, I present evidence for a positive response to this question.

## 3. Insincere Promises: an experimental study

### 3.1 Testing folk intuitions about lying

In the philosophical literature, it is generally agreed that a good definition of lying should track the ordinary usage of the term (Carson 2006:285; Fallis 2009:32): most debaters are after a characterisation of lying that is in line with the linguistic practice of competent speakers. A good account of lying should predict which usages of the term are correct and incorrect according to competent speakers. A corollary of this way of thinking is that if an account of lying makes predictions that are inconsistent with ordinary people’s intuitions, that account fails to meet one important *desideratum* of a theory of lying. With this in mind, philosophers and linguists have started to accumulate data about ordinary speakers’ intuitions about the correct usage of the terms. These studies have the potential to give us insight into what lying is, or at the very least into what lying is perceived to be within a community of speakers[[60]](#footnote-60).

Numerous and diverse empirical studies, stemming from different theoretical backgrounds and motivated by different explanatory aims, have attempted to explore folk intuitions about lying[[61]](#footnote-61). Among the ones explicitly investigating the intuitions of competent speakers about the concept of lying, the most important strand comes from the framework of prototype semantics. Following the lead of Coleman & Kay (1981), these studies attempt to identify the features of a prototypical lie, and to outline the differences of these prototypes across different cultures (Sweetster 1987, Cole 1996, Hardin 2010, Rong, Chunmei & Lin 2013). The present study addresses similar worries, but in a slightly different framework, namely that of experimental philosophy. Here, the aim of the analysis is to identify the necessary and sufficient conditions for an utterance to be a lie, rather than the prototypical features that make up the concept. Only a few studies on lying have been conducted within this framework so far, some attempting to test if the intention to deceive is necessary for lying (Arico & Fallis 2013, Meibauer 2016, Rutschmann, R., & Wiegmann, A. 2017), and others if actual falsity is (Turri & Turri 2015, Wiegmann et al. 2016). The present experimental study will instead try to establish which conditions are necessary for lying by promising.

### 3.2 Aim of the study

This experimental study aims to test the theories developed in section 2 against the intuitions of native English speakers. This means that it will attempt to determine if ordinary people rate illocutionary acts other than assertions as lies, and

which *insincerity conditions* have to be satisfied for them to do so. More specifically, this study will be concerned with one speech act in particular, namely promises. The main reason is that promises clearly display all of the speech-act theoretic puzzles introduced in chapter III: they can be performed by means of an explicit performative, they are not assertions, and their insincerity conditions are sensitive to attitudes other than belief (namely intentions). As a reminder, in the previous section I have introduced three candidate insincerity conditions for promises: BIC IIC or EIC:

**Candidate insincerity conditions for promising**

IIC: the speaker does not intend to Φ

BIC: the speaker does not believe that he will Φ

EIC: either the speaker does not believe that he will Φ, or does not intend to Φ, or both

Testing that illocutionary acts other than assertions are judged to be lies is relatively simple: it is sufficient to create a story in which it seems that a character lies by promising, and ask participants whether the character has lied. If participants classify promises as lies, we have evidence that not only assertions are classified as lies. Testing which insincerity condition is more accurate is a slightly more complex matter. Here we need different stories in which different combinations of insincerity conditions are violated, and for each story test whether the participants believe that the speaker has lied. Given our candidate insincerity conditions, we will need to consider three scenarios in which a character promises something insincerely. In the *straightforward scenario*, the character’s (S) utterance satisfies both BIC and IIC. In the *no-intention* scenario, it satisfies IIC but not BIC. In the *no-belief* scenario, it satisfies BIC but not IIC. I will refer to the latter two cases as the *crucial* conditions, as opposed to the *control* (straightforward) conditions.

• *Straightforward scenario*: BIC & IIC [Control]

• *No-intention scenario*: IIC & ¬BIC [Crucial]

• *No-belief scenario:* BIC & ¬IIC [Crucial]

### 3.3 The predictions of existing theories

In section 2 I have mentioned five approaches to lying by explicit promising. Each account gives different predictions about which of the three scenarios will be rated as a lie.

1. According to the *only-assertion paradigm* (∅), lying requires a direct assertion. Its prediction is that since promises are not direct assertions, respondents will claim that in no scenario is the character lying. Interestingly, if this view is correct, the first two speech-act theoretic puzzles are not really a worry for defining lying: since only direct assertions can count as lies, there is no issue of determining whether explicit performatives can count as lies, and under which conditions.
2. The *only-belief paradigm* (BIC) rigidly assumes that you lie only if you *believe* that the propositional content of your speech act is false. This view expects positive responses in the *straightforward* and in the *no-belief* condition, but negative responses in the *no-intention* condition. If this view is correct, the third speech act theoretic puzzle about insincerity is not really a worry for defining lying: since only beliefs are relevant to determine whether an utterance is a lie, there is no issue of extending the insincerity conditions to other attitudes.
3. The *only-intention paradigm* (IIC) maintains that a promise is insincere iff the speaker does not intend to fulfil his promise. Applied to lying, this view predicts that a promise is a lie iff the speaker does not intend to fulfil it. The *straightforward* and *no-intention* cases should then be rated as lies, but not the *no-belief* case.
4. According to a *cognitivist interpretation* (BIC & IIC), the *no-intention* and *no-belief* should be described as cases of irrational thinking rather than lying, so that only the *straightforward* case should be rated as a lie.
5. According to the *entailed-assertion* paradigm (BIC ⋁ IIC), a promise is a lie either if the speaker does not intend to fulfil his obligation, or if he believes that he will fail to fulfil it, or both. The account also predicts that all scenarios will be rated as lies, but expects the *straightforward* one to receive slightly higher ratings than the *crucial* ones.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Straightforward** | **No Intention** | **NO BEL** |
| **Only-assertion** | NO | NO | NO |
| **Only-belief** | YES | NO | YES |
| **Only-intention** | YES | YES | NO |
| **Cognitivist** | YES | NO (IRR) | NO (IRR) |
| **Entailed-assertion** | YES | YES (-) | YES (-) |

**Table 1**: the predictions of the five different accounts of lying by promising

### 3.4 Experiment 1

#### 3.4.1 Method

**Participants:** Participants were recruited using Amazon Mechanical Turk and tested using Qualtrics. They were compensated $0.2 for taking the survey. Repeated participation was prevented. Overall, 166 U.S. residents were tested (85 females; mean age (SD) = 36.6 years (12.9); range: 18–72; 100% reporting English as a native language). To prevent participants from taking the test negligently, a minimum response time (25 seconds) and a control question were set. Data from three participants who failed to meet these conditions were excluded, but including them would not affect the results.

**Design**: Each participant was randomly assigned to one of four conditions. Each condition features Coco and Baba, and in each condition Coco promises something to Baba. The first two conditions [(1) *straightforward lie*; (2) *no intention*] belong to the ‘drink’ story, and the second two conditions [(3) *straightforward lie*; (4) *no belief*] belong to the ‘repair’ story. In the ‘drink’ story, Coco promises not to drink; in the ‘repair’ story, Coco promises to repair Baba's car. For each pair, in the *straightforward* condition Coco lacks both intention and belief, and in the crucial case he lacks one attitude (intention in 2, belief in 4) but not the other.

Having been assigned to one of the conditions and having read the relevant story, the participants were posed two questions, always in the same order. The first asked whether Coco told a lie (“Did Coco tell a lie”? Y/N). The second allowed participants to report whether they felt uneasy in answering the preceding question (“Did you find it easy to make a decision”? Y/N); those answering “no” were solicited to motivate their uneasiness via a simple feedback form. The second question was especially designed to rule out the possibility that participants would have preferred not to give a dichotomic yes-no response, but also to collect qualitative data about the strength of the participants' intuitions.

Some peculiarities of the design are due to consistency constraints on the rationality of “intending without believing”, and vice versa (cf. Appendix II). The first peculiarity is that in all conditions Coco has a partial rather than outright belief in whether he will fulfil his promise. This is due to the adoption of *weak cognitivism* in this research, according to which intending to *Φ* entails not being certain that you will fail to *Φ*. To grant uniformity between all conditions, Coco has a partial belief both in the crucial cases (where it could not be otherwise) and in the straightforward ones. Similar consistency constraints (also discussed in Appendix II) motivated one asymmetry in the experimental design, *i.e.* the fact that the no-intention and no-belief cases were not tested within the same story. The reason is that an uncontroversial *no-intention* case demands a promise about *refraining from acting,* while an uncontroversial *no-belief* case demands a promise about *actively performing one action*.

#### 3.4.2 Results and discussion

Virtually every respondent rated the *straightforward cases* (in which Coco lacks both belief and intention to perform) as lies: 95% of the participants claimed that Coco lied in the *drink-straightforward* condition (38 of 40) and 95% in the *repair-straightforward* condition (39 of 41). All except one participant (in the drink scenario) declared that the question was easy to answer. The results for the straightforward cases support the view that it is possible to lie by explicit promising, refuting the *only assertion* hypothesis. They also back the stronger claim that insincere promises can be regarded as *prototypical* cases of lying; and that, more generally, a prototypical lie can be performed by uttering a speech act other than assertion.



**Figure 1**: Percentage of respondents rating the protagonist’s utterance as a lie in each condition.

In the *no-intention* condition, 90% of the participants (40 of 44) rated the promise as a lie. This refutes the predictions of the *only-belief* and of the *cognitivist* accounts, as respondents classed the promise as a lie even if Coco sincerely believes that he will (almost certainly) fulfil his promise. The difference between this condition and the corresponding *drink-straightforward* case was not significant, χ2(1, *N* = 84) = .53, *p* = .467; nonetheless, as much as 14% of the participants (6 of 44) declared that the question was not easy to answer, which is significantly more than in the *straightforward* conditions (Fisher's Exact Test, *N* = 125, *p* = .008, two tails)**.** This shows that intuitions in this condition are not as strong as in the *straightforward* cases, and suggests that the *no-intention* condition is regarded by some participants as a non-paradigmatic instance of lying.

The *no-belief* condition was rated as a lie by 73% of the participants (30 of 41). A binomial test showed that this result is significantly different to a chance distribution of the responses (*p* = .004**,** two tails); in other words, participants were more likely to say that the promise was a lie than the opposite (OR = 2.73), which logically entails that the *only-intention* account (according to which this case does not qualify as lying) can also be rejected. In this condition, 15% of the participants (6 of 41) declared that the question was difficult to answer: Fisher’s Exact Test revealed that this was significantly different from the straightforward cases (*p* = .006, two tails, OR = 13.71), suggesting that at least some participants did not see this as a paradigmatic case of lying. That the case might not be seen as prototypical is also confirmed by the lower percentage of ratings of the promise as a lie: this case differs significantly from all the rest of the cases jointly, χ2(1, *N* = 166) = 12.71, *p* = .001, as well as from the no-intention case separately, χ2(1, *N* = 85) = 4.6, *p* = .032.

****

**Figure 2***:* Percentage of respondents rating the question as difficult to answer (scale from 0% to 50%).

Interestingly, none of the feedback forms filled in by the participants who found it difficult to answer the question mentions the contrast between intention and belief as problematic, suggesting that the *cognitivist account* of intentions does not reflect the participants’ intuition. Logistic regressionwas used to test whether information about gender or age could be used to predict the responses, revealing that neither gender (*p* = .592) nor age (*p* = .808) was a significant predictor.

The results provide strong support for the definition of lying by promising elaborated in part A. However, one can still worry that the dichotomous yes/no design led some participants to polarize their responses in an unnatural way (cf. Xu, Lu, Fuo & Lee 2009:318, Turri & Turri 2015:163). In section 3 we have seen that beliefs, and hence insincerity, can come in degrees, and that contrasting parameters can influence our evaluation of a statement as insincere. Perhaps participants see lying as a graded or vague concept (cf. IV.4 and Bazzanella 2009) and would have preferred to give a graded answer. In fact, 14% of the participants in the crucial cases (12 of 85) declared that the question was difficult to answer. Maybe they gave a positive answer to sanction that the speaker misbehaved in some important way – but they would not have classed the utterance as a lie if an intermediate alternative was given. If this line of reasoning is correct, and the probes forced the participants to polarise their responses, the proposed interpretation might be mistaken. To test this hypothesis, the test was rerun with a slightly modified design that allowed for graded rather than dichotomous responses.

### 3.5 Experiment 2

#### 3.5.1 Method

**Participants:** Participants were 104 adults (57 females; mean age (SD) = 37.1 years (11.8); range: 19-64; 100% reporting English as a native language, recruited and tested online as in Experiment 1). One participant failing to meet the minimum response time was excluded, but including him would have not affected the results.

**Design:** The design was the same as Experiment 1, but instead of using a dichotomous yes/no measure, participants were asked to rate their agreement with the statement “Coco told Baba a lie” on a 7-point Likert scale (1: “definitely agree”, 4: “neither agree nor disagree”, 7: “definitely disagree”). If people had the intuition that the crucial conditions do not fully qualify as cases of lying, their ratings should concentrate around the midpoint (or below); if instead they shared the intuition that Coco is actually lying, the ratings should instead concentrate on the ‘lie’ end of the scale (5 or above).

#### 3.5.2 Results and discussion

In the *no-intention condition*, participants overwhelmingly agreed that Coco did lie: 84% (21 out of 25) of them rated it as “5” or above, and only 16% (4 participants) as “4” or below (mean = 6.0; mode = “7”). Similarly, in the *no-belief condition*, participants overwhelmingly agreed that Coco did lie, 88% (22 out of 25) of them rating it as “5” or above and the remaining 12% (3 participants) rating it below “4” (mean = 5.8; mode = “7”). Like in the first experiment, the mean score in the ‘repair belief’ condition was slightly lower than in the ‘drink intention’ condition, but the difference (unlike in Experiment 1) was not significant, *t*(48) = 0.64, *p* = .523. A comparison of the crucial and the straightforward cases revealed a significant difference between the two ‘repair’ scenarios, *t*(34.3) = 2.52, *p* = .017 (equal variances not assumed), but no significant difference between the two ‘drink’ scenarios, *t*(49) = 0.58, *p* = .562[[62]](#footnote-62). Overall, these results are well in accordance with those obtained with the dichotomous design, strengthening the case for the *entailed-assertion* account.

\\stfdata08\home\PI\Pip13nm\ManW7\Downloads\Pie chart(1).tif

**Figure 3:** Pie charts showing the participants’ ratings in each of the four scenarios.

A last worry to address is that participants might have agreed that Coco lied only because they were not allowed to categorise his statement in some alternative way that they found more adequate. Perhaps most had the intuition that in the crucial conditions Coco was deceptive, or insincere, and were led to describe him as lying only because no other category of assessment was offered. They agreed that Coco was lying because it was the only available option to express that he misbehaved, but they would have denied it if some alternative category more adequate to describe the situation, like being deceptive, was also available. To address this worry, a second study was conducted that provided participants with an opportunity to describe the speaker along two different categories: as being deceptive and as being lying[[63]](#footnote-63).

### 3.6 Experiment 3

#### 3.6.1 Method

**Participants:** Fifty-five new participants were tested (31 females, mean age (SD) = 34.6 years (12.5); range: 19–66; 100% reporting English as a native language, recruited and tested online as in the other experiments). Data from six participants failing to meet the minimum response time and/or the control question was excluded, but including them would have not affected the results.

**Design:** Participants were randomly assigned to one of the two crucial conditions (*no-belief, no-intention*) from Experiment 1. They were then asked to answer two questions that appeared (in randomized order) on the same screen:

* Did Coco say something deceptive? [Y/N]
* Did Coco tell Baba a lie? [Y/N]

Participants were then asked to answer the same control and demographic questions as in Experiment 1.

In both of the vignettes used in the experiment, there is no question that Coco’s statement is deceptive: in the *no-belief* condition he pretends to have a belief he does not have; in the *no-intention* one, he pretends to have an intention that he does not have. Participants who had the intuition that Coco was *not lying* thus had the opportunity to deny that Coco is lying while being able to describe Coco as misbehaving, namely as being deceptive.

Given that in both vignettes it is uncontroversial that Coco’s statement is deceptive, it is only the second question (about lying) that is of interest here. If the scores obtained in this experiment are significantly lower than those obtained in the same scenarios of Experiment 1, then the proposed interpretation of the results might be unwarranted. By contrast, if participants continue to describe Coco as lying, there is even stronger experimental evidence in favour of the proposed view.

#### 3.6.2. Results

The overwhelming majority of participants marked both the *no-intention* case (92%, 24 out of 26) and the *no-belief* case (90%, 26 out of 29) as a lie. The scores obtained in the crucial conditions in Experiment 3 are even higher than those obtained in Experiment 1, but neither of the across-experiment differences was significant (*no-intention*: χ2(1, *N* = 70) = 2.89, *p* = .089.; *no-belief*: χ2(1, *N* = 70) = .04, *p* = .84). These results are consistent with the previous ones, and strengthen the case for the *entailed-assertion* view. Somewhat surprisingly, the difference between the *no-belief* and the *no-intention* condition found in the previous experiments has disappeared, χ2(1, *N* = 55) = .12, *p* = .73.

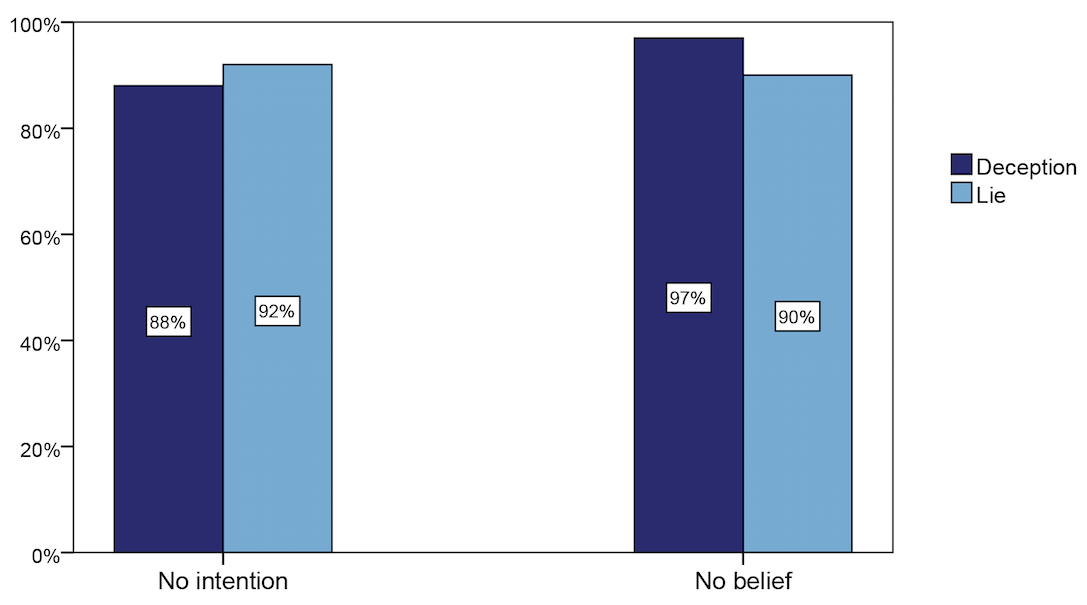


Figure 3**:** Percentage of respondents rating the protagonist’s utterance as a lie and as deceptive.

### 3.7 General discussion

The results of all experiments strongly support the *entailed-assertion* account. They consistently show that it offers the best predictions of people’s intuitions about lying by promising. A promise is a lie iff the speaker lacks belief or intention to fulfil the promise, or both. Furthermore, these results undermine all alternative views: the *only-assertion* view that only assertions can count as lies, the *only-belief* view that believing that what you say is false is necessary for lying, and the *only-intention* view that a promise is a lie only if the speaker does not intend to perform. In doing so, they also confirm the significance of the speech-act theoretic puzzles about lying, as these puzzles are directed at criticising both the *only-assertion* and *only-belief* view about lying.

In Experiment 1 and 2, the crucial conditions obtained lower scores than the control conditions (Experiment 3 had no control conditions), and in Experiment 1 around one in six participants declared that they found it difficult to decide whether to classify them as lies. In line with the prediction of the *entailed-assertion* view, this suggests that the consensus is less pronounced when only one of the two sincerity conditions for promising is violated, and that a portion of the population sees them as non-paradigmatic instances of lying.

### 3.8. More on the results

#### 3.8.1 Intentions vs Beliefs

The crucial conditions got different results in Experiment 1: the *no-intention* condition got significantly higher rates than the *no-belief* one. A slighter difference between them was also found in Experiment 2. This kind of difference, however, was not found in Experiment 3, or in the second question of Experiment 1 (about the participants’ uneasiness to define Coco’s utterance as a lie). Perhaps the difference is due to random fluctuations in the subjects’ intuitions, and does not require an explanation. But if an explanation must be given, it can be given in terms of *moral judgements*.

Several authors have suggested that lying is a morally loaded term (Bok 1978:14, Williams 1985:140), and some of them even contended that white lies are not lies (Margolis 1962, Donagan 1977:89, Grotius, RWP:1212-8). A plausible view is that moral judgements might affect whether one finds a particular case of lying more or less prototypical. Experimental studies have also shown that judgements about intentions (Knobe, 2003) and causation (Alicke 2014) can be influenced by judgements about culpability (*i.e.* moral judgements). If moral judgements can affect folk intuitions about whether a particular case is a lie, a moral asymmetry between the scenarios might have influenced the results. In fact, there is such an asymmetry between the *no-intention* and the *no-belief* cases. In the first case, Coco is fully responsible for not fulfilling the promise: it is in his power to fulfil it, but he willingly decides to infringe it. By contrast, in the no-belief case Coco intends to do what is in his power to fulfil the promise, but it is not fully in his power to do so. In other words, in the first case he is fully responsible for the infraction, while in the second he is only responsible for having set the stakes too high.

While the no-intention case is by definition in ‘bad faith’, the no-belief case is by definition in ‘good faith’. This could explain the slight asymmetry in the results – an asymmetry also expected in any replication of the test, given that it is built into the difference between violating the ‘belief sincerity condition’ and violating the ‘intention sincerity condition’.

#### 3.8.2 The intention to deceive condition

Proponents of the *intention to deceive condition* (IDC) for lying (cf. chapter II) might be worried that these results have been influenced by the fact that it has been left unspecified whether Coco intends to deceive Baba. For instance, the *no-belief* case might have received lower results in Experiment 1 because it is not clear whether Coco intends to deceive Baba. In this section, I will show not only that this worry is unfounded, but that the results undermine the very idea that ordinary speakers take the IDC to be necessary for lying.

On a standard interpretation (the “rigid” intention to deceive condition, or IDC2, cf. II.2.2), the relevant intention to deceive has to be about the content of the statement: if the content of the statement is *p,* the speaker has to intend to make the hearer believe that *p.* On a weaker version of the IDC (the “believed sincerity condition”, or BSC, cf. II.2.3), the relevant intention is just to make the hearer believe that *the speaker believes* that *p.* In other words, where *p* is the *believed-false* content of the speaker’s statement, the two most influential versions of IDC are:

IDC2: S intends A to believe: (p)

BSC: S intends A to believe: (S believes that p)

In the *no-belief* condition, the content of the promise is that Coco will repair the car, a proposition that Coco believes to be very likely false. Now, unless Coco intends his promise not to be accepted, or not to be acted upon (*i.e*. if Coco’s promise is a normal promise), Coco clearly intends Baba to believe that he will repair the car, so the participants have no reason to believe that IDC does not obtain.

In the *no-intention* condition, by contrast, the relevant content is that Coco will not drink. But here Coco believes the proposition to be probably true, so that neither IDC2 nor BSC can obtain, and the participants cannot think that they obtain. Two consequences can be drawn from this observation. The first one is that intuitions about the intention to deceive condition did not alter the result of the experiment: if they had some weight, they would have favoured the *no-belief* case over the *no-intention* one; instead, it was the latter that obtained significantly higher results. The second is that the high results from the *no-intention* case (90% lies, showing no significant difference from the straightforward cases) strongly suggest that neither IDC2 nor BSC is perceived as a necessary condition for lying.

Here is a reply. Even if neither IDC1 nor BSC are satisfied in the *no-intention* case, Coco is still aiming to deceive, since he clearly intends Baba to believe that he intends not to drink. The problem with this response is that it relies on a problematic definition of ‘intention to deceive’ (paralleling the “broad” intention to deceive condition, or IDC1, from II.2.1) according to which there are no constraints on what the deception is about:

IDC1: S intends to deceive A

The problem with the IDC1 is that we have already seen that it is untenable, because it counts any deceptive believed-false statement as a lie, even if deception has nothing to do with the content of the statement. It seems that the IDC should instead only capture deceptive intents that are somehow related to what is said by the liar. As a matter of fact, there are known counterexample to IDC1, as the Theatrical Pose counterexample by Fallis (2010:6) (cf. II.2.1).

To be able to defend the claim that Coco is lying in the *no-belief* case, the proponent of the IDC has to provide a different version of the IDC – perhaps one that is sensitive to the different attitudes expressed by a speech act, such as IDC4:

IDC4: S intends A to believe that (S Ψ(p)),

where Ψ is the propositional attitude expressed by the illocutionary act with content *p* that S has performed.

To conclude, the experiment represents a challenge to the existing versions of the IDC, and suggests a further challenge for their proponents: that of constraining the content of attempted deception in a way that generalises across different illocutionary acts.

#### 3.8.3 Falsity condition

In the literature on lying, virtually every author accepts a ‘*subjective’* account of lying, according to which asserting an *objectively false* proposition is not necessary for lying, as long as the speaker *believes* that proposition to be false (cf. 1.1.2). However, a few authors (Grotius RWP:1209, Benton forth) have suggested that falsity of the statement, in addition to belief in its falsity, is required for lying – call this the *objective* view. Recently, Turri & Turri (2015) claimed to have found experimental evidence that most laypeople endorse the objective view. However, their study is far from convincing, and has recently been dismissed by Wiegmann et al. (2016). An interesting aspect of the present study is that it puts further pressure on the objective view of lying, and on Turri & Turri’s claim that such view reflects laypeople’s intuitions.

How does the present study relate to this debate? A first suggestion is that if participants regarded falsity as necessary for lying, they would not have rated a promise as a lie unless the promise was actually infringed in the story. However, in all conditions of all experiments participants agreed that the protagonist lied, even if it is always left unspecified whether the promise was fulfilled or not, *i.e.* if it was objectively false[[64]](#footnote-64). This seems incompatible with the fact that all respondents in the straightforward conditions, and the majority in the others, indicated that Coco lied. Moreover, in the straightforward cases, 99% of the participants reported that they found it easy to respond. How could this be, if they did not know whether the falsity condition obtained?

One easy response is that in all four cases the respondents predicted that, given the information available, the promise would eventually be infringed in the story: they took it that it was ‘implicit’ that the falsity condition would obtain. This is clearly plausible for the *straightforward* conditions: since Coco intends not to do what he promised, and believes that very probably he will succeed in not doing it, the falsity condition will almost surely be met. A similar inference is plausible in the *no-belief* condition: even if Coco intends to repair the car, he believes that very likely he will not succeed, and this clearly suggests that the car will not be repaired.

The real problem for proponents of the falsity condition is the *no-intention* case: here, Coco intends to drink against his promise, but he believes that he will very likely fail to do so, because he will not be able to. The information provided in the scenario cannot license the inference to the conclusion that he will drink; as a matter of fact, it only licenses the opposite inference. In other words, not only it is not specified if the falsity condition obtains, but the scenario clearly suggests that it will not obtain. In all experiments, the *no-intention* condition was consistently rated as a lie (obtaining even higher scores than the other crucial condition), and in no experiment its results were significantly different to the straightforward scenarios: this strongly suggests that participants did not take falsity of the promise to be necessary for lying.

The experimental data collected thus strengthens the case for the subjective account. As already shown by Wiegmann et al., people think that lying does not require falsity after all[[65]](#footnote-65).

### 3.9 Conclusions

In the previous chapter, I have developed a speech-act theoretic framework to outline the necessary and sufficient conditions for lying by promising, and sketched a way to extend this account to other illocutionary acts performed by uttering a sentence containing an explicit performative. Assuming that an ‘assertion-based’ definition of lying is broadly correct, the proposed account takes a promise to be a lie under the following conditions:

**Definition of lying by promising**

In successfully uttering a promise with content p, S lies to A about p iff:

1. S thereby asserts that p
2. Either S believes that not p, or S does not intend to p, or both

More generally,

**Speech-act theoretic definition of lying**

In successfully uttering an illocutionary act with content p that expresses an attitude Ψ(p), S lies to A about p iff:

1. S thereby asserts that p, i.e.:
2. S expresses p
3. S presents p as an actual state of affairs
4. S takes responsibility for p being an actual state of affairs
5. Either S believes that not p, or S is not in Ψ(p), or both

In introducing my speech act-theoretic definition of lying, I have put into question several ideas that are found in the literature: the view that only direct assertions can be lies; the view that the propositional content of a lie has to be believed to be false; the view a promise is insincere only if the speaker does not intend to perform; the view that falsity is required for lying. My experimental study has shown that, at least with respect to promises, the proposed view (but none of the alternative ones) gives predictions that are consistent with folk intuitions about lying.

Overall, it seems that my proposed definition offers a plausible account of the conditions under which the performance of an insincere speech act counts as a lying, that reflects ordinary people’s judgements about whether a given utterance is a lie.

#### Appendix I – the scenarios

Drink story - Refraining from action

Baba and Coco are married. Coco is going to a party, but Baba does not like Coco to drink. So Baba asks him: “Will you be drinking alcohol at the party?”. Coco replies: “Do not worry Baba: I promise that I will not drink alcohol at the party.”

*Scenario 1 - Straightforward:*

In fact, Coco intends to drink alcohol at the party, and he is almost certain that he will find something to drink there.

*Scenario 2 - No intention (to refrain as promised):*

Coco actually intends to drink alcohol at the party and he will attempt to, but he is almost certain that he won’t succeed, since he believes that the hosts do not offer alcoholic drinks at their parties.

Repair story - Positive action

Baba has broken her car, but she needs it to visit her family next week. For this reason, Baba has called Coco the mechanic to repair it. Coco the mechanic checks the car and tells Baba: “Do not worry Baba: I promise that I will repair your car by next week.”

*Scenario 3 - Straightforward:*

Coco has no intention whatsoever to repair the car, and he is almost certain that he will not repair it.

*Scenario 4 - No belief:*

Coco intends to repair the car and he will attempt to do it, but he is almost certain that he won’t manage to repair it in the end, because the damage is too serious.

## 4. Gradedness

The previous sections have extended the insincerity conditions for lying to attitudes other than assertions. This is an important advance in the definition of lying, but it still overlooks one often-undervalued feature of insincerity: the fact that it comes in degrees. As Montaigne nicely stated, while truth is unique, “the opposite of truth has many shapes, and an indefinite field” (Montaigne *E:* 1.IX). There is a whole grey area of ‘half-sincere’ utterances that are difficult to classify and, quite importantly, it is in this grey zone that liars strive.

To shed some light on this obscure area, this section will consider cases involving partial insincerity; for example, statements that are not fully believed to be false, but that are nevertheless not believed to be true.Are these statements lies? And how much confidence in their falsity is required for them to count as lies? We will discuss such questions, and explore the thin, elusive line that distinguishes a sincere assertion from an insincere one. This will be a hard challenge, and indeed for a theory of lying “the more difficult task [is] that of drawing lines” (Bok 1989:49).

The standard, simplistic account of the insincerity condition for lying is that an utterance is a lie only if the speaker believes it to be false. However, the expression “believe to be false” is not really helpful to deal with intermediate cases, as it does not specify what degree of confidence in the falsity of *p* counts as believing it be false. In what follows, I will develop a version of the insincerity condition that be able to classify statements that are neither fully believed to be false, nor fully believed to be true. For most of my discussion, I will pretend to adopt the *only belief* view: this will allow me to provisionally set aside the intricacies that arise when we consider attitudes other than belief. In the last section, I will extend the analysis to other attitudes, to provide a final, complete account of the insincerity conditions for lying.

### 4.1 The dichotomic view and the traditional insincerity condition

Both deceptionist and non-deceptionist typically agree on the following formulation of the insincerity condition for lying (call it the ‘traditional insincerity condition’, or TIC):

(TIC) S believes that *p* is false

Scholars endorsing this condition tend to take for granted that a statement is *sincere* when the speaker believes it to be *true,* and *insincere* when the speaker believes it to be *false*[[66]](#footnote-66), and that a more fine-grained analysis would be unnecessarily intricate (Saul 2012:5, fn10). From this perspective (call it the *dichotomic view*), the definition of lying correctly rules out only statements that are believed to be true.

The Dichotomic View

A statement is *sincere* when the speaker believes it to be *true,* and *insincere* when the speaker believes it to be *false*, *tertium non datur*

It is not obvious, however, that the dichotomic view is correct, nor that TIC offers a satisfying characterisation of the insincerity condition for lying (Mahon 2015:1.5). There is good ground to suspect that the dichotomic view is not an adequate assumption defining lying, because a number of intermediate credal states[[67]](#footnote-67) exist between believing *p* to be true and believing *p* to be false.

First, it is possible for a speaker to believe that a statement is only partly false (rather than utterly false): in this case, the speaker believes that *p* has a ‘graded truth value’. Second, it is possible for a speaker not to be certain; in other words, to have a graded degree of confidence (rather than a flat-out belief) in the falsity of a statement: intermediate beliefs of this kind are called ‘graded beliefs’. The difference between these two layers of gradedness can be difficult to grasp: in section 4.2 and 4.3 I will explain in detail this subtle distinction.

In what follows, I introduce two counterexamples to the dichotomic view: namely, lies that involve beliefs about graded truth values (4.2) and lies that involve graded beliefs (4.3). I develop a non-dichotomic alternative to the TIC that counts this kind of statement as lies and allows for various degrees of insincerity in lying, according to which the speaker must believe his statement to be *more likely to be false than true* (4.4).

### 4.2 Graded Truth Values

Every species is vague, every term goes cloudy at its edges, and so in my way of thinking, relentless logic is only another name for stupidity—for a sort of intellectual pigheadedness.

H.G. Wells, First and Last Things (1908)

A first motive to challenge the dichotomic view emerges if one considers the question of lying from outside the framework of bivalent (*i.e.* two-valued) logics. According to traditional, bivalent logics, the truth value of a proposition is either true or false, *tertium non datur*, so that there is no point in distinguishing between a statement that is false and a statement that is not true. By contrast, many-valued logics allow for a larger set of truth values. If ‘true’ and ‘false’ are not the only two possible truth-values that one can assign to propositions, the assumption that speakers can only have beliefs involving these two truth-values strikes as spurious, or at least unwarranted.

In the literature on lying, Chisholm & Feehan have offered a discussion of insincerity that can be read as a challenge to a bivalent account of insincerity[[68]](#footnote-68). According to their alternative insincerity condition, a speaker lies only if he states what he believes to be *false* or *not true* (I call this the *fuzzy insincerity condition*, FIC).

**Fuzzy Insincerity Condition**

FIC: S believes p to be false or not true

This alternative formulation of the insincerity condition relies on the subtle difference between believing that *p* is *false* and believing that *p* is *not true*. Chisholm & Feehan (1977, 152) note that “it is logically possible to believe one of these things [e.g. *p* is not true] and not the other [e.g. *p* is false]”. One way to interpret this claim is to consider the difference between false and not true within a specific many-valued logic theoretical framework, namely *fuzzy logic*[[69]](#footnote-69)*.*

Fuzzy logic is a many-valued logic conceived especially for predicates that are intrinsically vague (like being bald, or old, or happy) and that, being graded in nature, allow for a number of truth values. Fuzzy logic takes as truth values all real numbers between 0 and 1, where 0 is false and 1 is true. From this perspective, to say that the speaker believes that a proposition *p* is *not true* is to say that the speaker believes that the truth value of *p* is x, where x is 0≤x<1. By contrast, to say that the speaker believes that a proposition *p* is *false* is to say that the speaker believes that the truth value of *p* is 0. Against the dichotomic view, stating what is *believed to be not true* is thus not the same as stating what is *believed to be false*. Believed-false statements are a subset of insincere statements.

Let us call lies that involve these intermediate beliefs (beliefs about graded truth values) *fuzzy lies*. Now, consider an example of *fuzzy* *lie* to grasp the difference from traditional lies. Suppose that Mickey utters (1) to persuade Daisy to date Donald:

(1) Donald is not bald

If Mickey thinks that Donald is *almost definitely* bald (e.g. he believes that (1) has a truth value of 0.2), he does not say something that he believes to be utterly false, and therefore the TIC does not count (1) as lying. However, intuitively Mickey is lying[[70]](#footnote-70). The fuzzy insincerity condition FIC accommodates our intuitions in counting (1) as a *fuzzy* *lie*, since (1) is believed to be *not true* (0.2 is less than 1 and more than 0).

FIC is broader than TIC: it allows all statements whose truth value is believed to be less than 1 to count as lies. Moreover, it correctly rules out misleading statements (given that they are believed to be true), while ruling in all standard lies (given that they are believed to be false). Nevertheless, it seems patent that the FIC is too broad, for the set of statements it allows to count as lies is too large. For instance, if Mickey believes (1) to have a truth-value of 0.8, FIC would predict that Mickey is lying, but this is counterintuitive, as Mickey in this case believes that Donald is *almost definitely* not bald. To avoid this problem, one could narrow the FIC in order to require that the believed truth value of the statement be closer to falseness than truthfulness– make it *x*, where *x* is 0≤x≤0,5 (call this the *revised fuzzy insincerity condition*, henceforth FIC\*).

**Fuzzy insincerity condition, revised**

FIC\*: S believes that p has a truth value comprised between 0 and 0,5

This solution is nevertheless problematic, since there seems not to be a clear theoretical basis to set the limit at a precise value. If one accepts that a speaker who believes that the truth value of his statement is 0.5 is lying, then it also seems reasonable to accept that a speaker who believes that the truth value of his statement is 0.51 is lying. But the same line of reasoning would work for the successive values (0.52; 0.53; [...]; 1), so that, in the end, all statements would count as lies.

A further problem with this view concerns the very existence of such credal states: the proposed representation of the speaker’s beliefs is so fine-grained that it may appear to overrate reality. It seems quite clear that in real life we do not experience the threshold between believing a statement to have a truth value of 0.50 rather than 0.51. If such a threshold exists, it is not consciously perceived; and since lying is a conscious choice, no such threshold can be taken as a necessary condition for lying.

Rather than positing a sharp ‘numerical’ threshold that separates insincerity from sincerity, one could limit the FIC to require that the believed truth value of the statement is *perceivably* closer to falseness than truthfulness.

**Fuzzy insincerity condition, revised again**

FIC’: S perceives p’s truth value to be closer to falseness than truthfulness

The revised condition FIC’ may seem rough compared to its ‘numerical’ translation, but it acknowledges that our beliefs only roughly (and rarely) correspond to the subtle differences that fuzzy logic outlines. Whether or not one finds this revised definition convincing, eventually we will be forced to abandon it: as I will show in the next section, it incorrectly rules out lies that involve graded beliefs.

### 4.3 Graded beliefs

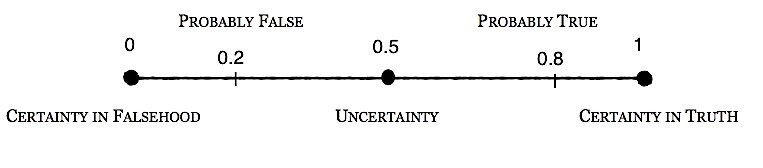
The dichotomic view holds that we either believe something to be true, or we believe something to be false. This is certainly true if we restrict our analysis to cases of certainty. Here, by certainty I am referring to what philosophers call ‘psychological’ certainty[[71]](#footnote-71): the highest degree of confidence that a subject can have in the truth of a proposition. As long as this state of mind is concerned, it is certainly true that one can only be supremely confident that a proposition is true, or supremely confident that it is false.

It does not seem, however, that we can only be supremely confident in the truth or falsity of a proposition. Quite the contrary: many of the beliefs we hold in our daily life involve a certain degree of uncertainty. This prompts the question of which kind of belief is involved in cases where certainty is not present. If certainty is the highest degree of confidence, there must be beliefs that involve a degree of confidence lower than certainty. These weaker beliefs, that do not fit within the framework of the dichotomic view, are known in the literature as ‘credences’, or ‘graded beliefs’.

That ordinary beliefs can be graded is evident if one thinks about daily situations in which a subject lacks certainty in a proposition and nonetheless, to some extent, believes that proposition. Consider some further examples: suppose Groucho holds the following three beliefs:

1. I have a pair of moustaches
2. Bulgaria will beat Azerbaijan in their next football match
3. There is life on some other planet in the universe.

At T1, Groucho regards (1) as certain, (2) as probable, (3) merely as more likely to be true than not. Groucho neither fully believes nor fully disbelieves (2) or (3). His partial beliefs in (2) is (3) (believing to be *probable*, believing to be *unlikely*, etc.) are what philosophers call ‘graded beliefs’, because they can be ordered in a graded scale[[72]](#footnote-72): Groucho is more confident in the truth of (1) than he is in (2), and in (2) than he is in (3). Formal accounts of degrees of belief (namely Bayesian accounts) represent this scale with real numbers from 0 to 1, where 0 indicates certainty in the falsity of *p*, 1 indicates certainty in the truth of *p*, and 0.5 indicates uncertainty – in other words, that the subject regards *p* just as likely to be true as false. On this view, uncertainty is the middle point (0.5) of a continuum of degrees of belief whose poles are certainty in the falsity (0) and in the truth (1) of the proposition (cf. Figure 1). To provide a formal account of the previous example, one could say that Groucho has a degree of belief of 1 in (2), of 0.75 in (3), of 0.55 in (4).



**Figure 1**: A visual representation of the certainty-uncertainty continuum

The fact that epistemic agents can hold a wide array of graded beliefs is at odds with the dichotomic view, that only allows for full belief in truth and full belief in falsity. Since graded beliefs and uncertainty are ordinary psychological states, it seems that a theory of lying should account for them (Meibauer 2014: 223, D’Agostini 2012:41). For instance, suppose that Groucho states (2) (that Bulgaria will beat Azerbaijan) while believing that it is probably false, or as likely to be false as true. Would his statement be sincere or insincere? More generally, how are we to draw the boundary between sincere and insincere utterances, and (consequently) between lies and not lies?

### 4.4 A ‘graded’ definition of insincerity

To see that the standard account of insincerity struggles to handle graded beliefs in a satisfactory way, let us consider a new example, inspired by recent historical events (cf. Carson 2010:212-21): George is a political leader, and tells (1) to a journalist. Propositions (a), (b), and (c) indicate George’s degree of confidence in his utterance, in four possible scenarios[[73]](#footnote-73).

1. Iraq has weapons of mass destruction
   1. (1/¬p) [Iraq has *certainly* *no* weapons of mass destruction]
   2. (0.75/p) [*Probably*, Iraq has weapons of mass destruction]
   3. (0.75/¬p) [*Probably*, Iraq *does not* have weapons of mass destruction]

Scenario (1a) is a clear-cut case of lying, since George believes (1) to be *certainly false:* the traditional insincerity condition (TIC) correctly tracks the intuition that, since George believes (1) to be false, (1) is a lie. In (1b), by contrast, George believes the statement to be *probably true:* even if he is not completely confident that the statement is true, it seems that in this case he is not lying (but cf. Marušić 2012:8). The utterance is inaccurate, and perhaps misleading, because it misrepresents George’s degree of belief in (1). However, being inaccurate or misleading is clearly not the same as lying (Saul 2012, Stokke 2013b). Also in this case, TIC makes the right predictions.

Problems arise for scenario (1c), where George believes (1) to be *probably false*. It seems that TIC does not count this case as a lie, because George does not utterly believe (1) to be false[[74]](#footnote-74). However, intuitively this is a case of lying, because George is saying something he believes to be very likely false. Since it excludes this sort of cases, TIC is too narrow, and needs some refinement.

Cases like (1b,c) suggests that a more fine-grained account of lying is needed, one that appreciates how lying can involve graded beliefs. The fuzzy insincerity condition (FIC’) will be of little help here. Such condition accounts for lies that involve *beliefs about graded truth values* (*fuzzy lies*), but it does not account for lies that involve *graded beliefs* (*graded-belief lies*) about plain truth values. Such a subtle difference is worth an explanation.

Let us represent the general structure of beliefs as “B(*p*)”, where the variable “B” takes beliefs as values, and the variable “*p*” takes the truth-value of the propositional content of beliefs as values. The dichotomic view assumes that both “*p*” and “B” can assume as values only 1 or 0: either a subject believes *p,* or he does not believe *p*, and either he believes *p* to be true, or he believes *p* to be false.

Non-dichotomic accounts, by contrast, assume that “*p*” and/or “B” can take as values all the real numbers from 0 to 1. *Fuzzy lies* involve non-whole “*p”* values, while *graded-belief lies* involve non-whole “B” values. In 4.2 I provided an example of a *fuzzy lie*; let us contrast it with the *graded-belief lie* now:

1. Iraq has weapons of mass destruction
2. Donald is not bald

If George is confident, but not sure that (1) is false (*e.g.*, a degree of confidence of 0.2 in (1)), it seems clear that George is lying: his lie is a *graded-belief lie*. In this case, 0.2 expresses the value of “B”, so that 0.2 indicates George’s subjective degree of confidence in the probability of (1). This case is different from the *fuzzy lie* example discussed in 4.3: in that case, 0.2 indicated the truth value of (2). In the fuzzy lie example, Mickey had an outright belief (B=1) that Donald is almost definitely bald – *i.e.,* that the truth value of (2)is 0.2.

The FIC allows for fuzzy lies like (2) to count as lies, but do not count graded-belief lies like (1) as lies, and is therefore too narrow. An alternative insincerity condition that allows *graded-belief lies* can be found in Carson (2006: 298). His proposal comes in two varieties: he presents a strong and a weak version of his ‘insincerity condition’ for lying. The first, ‘strong’ version requires that the speaker believe his assertion to be “false or probably false”. Let us call Carson’s first condition the ‘strong insincerity condition’ for lying (SIC):

**Strong insincerity condition**

(SIC)S believes p to be at least probably false[[75]](#footnote-75)

SIC correctly captures prototypical cases of lying like (1a) (repeated below). Unlike the traditional definition, it also includes lies that are not believed with certainty to be false, like (1c), that George believes to be *probably false*. This is an advantage of SIC over the traditional condition TIC, since it seems intuitive that saying what you believe to be *probably* false counts as lying – even if it is arguably less insincere, and less deceptive, than a full-fledged lie.

However, the limit set by the SIC strikes as arbitrary: it is not clear what justifies drawing the boundary between sincerity and insincerity exactly on the degree of confidence indicated by ‘probably’, and not someplace else. The term ‘probably’ indicates a degree of confidence in the proposition higher than uncertainty and lower than certainty: for the sake of the argument, let us assume it stands for a degree of belief of 0.75 or higher. If a degree of belief of 0.75 in the falsity of the proposition is enough for lying, there seems to be no reason to exclude lower graded beliefs like 0.7, or 0.6, that are perceivably higher than uncertainty (0.5).

1. Iraq has weapons of mass destruction
   1. (1/¬p) [Iraq has *certainly* *no* weapons of mass destruction]
   2. (0.75/¬p) [*Probably*, Iraq *does not* have weapons of mass destruction]
   3. (0.6/¬p) [*Presumably* Iraq *does not* have weapons of mass destruction]

For instance, in (1d), George utters what he believes to be *more likely to be false than true*, so that it seems that he is lying. However, SIC does not capture (1d), because by hypothesis George’s degree of confidence is higher than uncertainty but falls short of believing (1) to be probably false. In failing to account for the intuition that also (1d) is a lie (even if arguably less insincere than (1c)), SIC is too restrictive. Furthermore, it is not clear that Carson’s SIC is able to capture fuzzy lies: only outright falsity is mentioned in the formulation, leaving it unspecified whether believing a proposition to have a truth value of more than 0 would qualify to satisfy SIC.

Carson’s second, ‘weak’ proposal avoids both problems. The ‘weak insincerity condition*’* (WIC) posits that lying requires that the speaker “does not believe [the asserted proposition] to be true” (Carson 2006, cf. also Davidson 1985:88, Sorensen 2007:256, 2011:407, Fallis 2013:346).

**Weak insincerity condition**

(WIC) S does not believe p to be true

Since it acknowledges that utterances like (1d) are lies, WIC is preferable to SIC. Furthermore, WIC seems compatible with fuzzy lies. Arguably, if S believes *p* to have a truth-value inferior to 0.5 (or perceives that value to be closer to falseness than truthfulness), S does not believe that *p*. If we accept this principle, WIC is broadly equivalent to FIC’.

However, the WIC is too broad: it incorrectly captures cases in which the speaker has no idea whether what he says is true or false, but goes on saying it for some independent reasons. These cases are classified in the literature as bullshit (Frankfurt 1986). The typical example of bullshitter is the politician who “never yet considered whether any proposition were true or false, but whether it were convenient for the present minute or company to affirm or deny it” (Swift 1710). For instance, consider the following example of deceptive bullshitting. Nick is a politician who does not know what the acronym LGBT refers to. When asked by a journalist about his opinion on LGBT rights, Nick answers:

1. LGBT rights are of central importance for our party

In uttering (3), Nick does not have the slightest idea whether what he said is true or false. His only concern is to trick the journalist into thinking that he knows what he is talking about. It seems that he is not lying, but the WIC counts incorrectly his statement as a lie, since Nick does not believe that his statement is true. As a matter of fact, philosophers seem to agree that, as long as the speaker has no opinion about the veracity of what he is saying, his utterance is better classified as a misleading utterance than as a lie (Saul 2012:20, Meibauer 2014: 103, but cf. Falkenberg 1988:93, Carson 2010:61-2); if one wants to account for this intuition, the WIC is too broad.

Since the SIC is too narrow and the WIC is too broad, an ideal condition has to lie somewhere in the middle. To find a middle ground between these two proposals, we can require that the speaker believe *p more likely to be false than true*. Call this the *comparative insincerity condition:*

**Comparative insincerity condition**

CIC: S believes p more likely to be false than true

Unlike WIC, CIC correctly rules out bullshit and statements uttered in cases of uncertainty. Unlike TIC, it counts graded-belief lies as lies. And unlike SIC, it rules in the other cases in which the speaker does not believe the statement to be true – like (1c) and (1d).

A first worry about the CIC is that it implicitly accepts the view that every belief can be represented as an assessment of probability. In case one finds this hypothesis disputable, one might prefer a phrasing that avoids a terminology that is committed to this view. Furthermore, it might be argued that it is not clear that CIC rules in fuzzy lies. We might stipulate, in a similar spirit of what we did for WIC, that if S believes *p* to have a truth-value inferior to 0.5 (or perceives that value to be closer to falseness than truthfulness), S thereby believes *p* more likely to be false than true. But an even better solution might be to introduce a phrasing that avoids both these worries:

**Comparative insincerity condition, revised**

CIC\*: S is more confident in ¬p than he is confident in p

CIC\* has the same strength points of CIC, but on top of that it clearly rules in *fuzzy lies,* and is not committed to understanding graded beliefs in terms of assessment of probabilities. A last worry might survive about the very assumption that there is a clear-cut boundary between insincerity and sincerity. Perhaps there are indeterminate cases that amount to neither lying nor not lying, and we should treat insincerity and lying are vague predicates – a similar problem, after all, seemed to apply to FIC. But this intuition can be accommodated without altogether rejecting CIC: unlike other insincerity conditions, this is the only one that allows for a progressive transition from sincerity to insincerity. On the other hand, if lying and insincerity are not vague predicates and a neat point of transition is to be individuated, the CIC is fine-grained enough to identify the boundary that gets closer to our intuitions, avoiding the counterexamples to which the alternative accounts fall victim (and if one has worries similar to those that applied to FIC, CIC\* can be qualified by requiring that the speaker is *perceivably* more confident in ¬p than he is confident in *p*).

### 4.5 Expressing graded beliefs and graded truth values

We have seen that an assertion is insincere if there is a certain discrepancy (defined by CIC\*) between *the speaker’s belief* (henceforth BΨ) and *the belief expressed by the sentence* (henceforth BΛ). This discrepancy can come in degrees, because of the graded nature of beliefs and of the content of beliefs – *i.e.* the graded nature of BΨ discussed so far. For a complete picture, we need to look at the other side of the coin: how insincerity is affected by the different degrees of belief that an assertion can express – the graded nature of BΛ.

Speakers employ several linguistic devices to express graded beliefs, or belief about graded truth values. We know from our experience as ordinary language speakers that it is possible to modulate the intensity of a statement, either mitigating or reinforcing it, thereby altering the strength of the belief expressed in BΛ. For instance, instead of simply uttering (1), a speaker can alternatively downgrade his assertion by uttering (1\*) or emphasise it by uttering (1\*\*):

1. Giusi is pretty

(1\*) Giusi is *kind of* pretty *somehow*

(1\*\*) *Believe me,* Giusi is *absolutely* pretty

Now, the previous section has considered graded-belief lies, in which BΨ is graded. Graded assertions like (1\*) and (1\*\*), by contrast, are cases in which BΛ is graded: the former is intuitively a weaker assertion, whereas the latter is intuitively a stronger one. In pragmatics, the two opposite phenomena of mitigation (1\*) and reinforcement (1\*\*) have often been studied separately, and labelled with different names: for the former, “attenuation”, “weakening” and “downgrading”; for the latter, “strengthening” and “emphasising” (see Fraser 1980, Coates 1987, Bazzanella et al. 1991, Caffi 2007, Egan & Weatherson 2011). The label of intensity (Holmes 1984, Labov 1984) unifies these two opposite directions of modulation.

The intensity of an utterance can be modified along different dimensions. In what follows, I discuss how intensity markers can modify the *propositional content* of an assertion to express different graded truth-values (4.5.1); and modify the *illocutionary force* of an assertion, to express different graded beliefs (4.5.2-3).

#### 4.5.1 Intensity and propositional content

The *propositional content* of a statement can be modulated both on the axis of quality (precision) and of quantity (augmentation or diminution). Expressions like “a little”, “very”, “much”, or “quite” are used to modify intensity on the axis of quantity. These linguistic devices allow the speaker to slightly alter the truth conditions of his statements (Lakoff 1973, 478-488). For instance, if Bruce utters (2\*) rather than (2), he quantifies Robin’s gladness to a lower degree, thus altering the truth-conditions of his statement:

(2) Robin is glad

(2\*) Robin is *pretty* glad

In section 4.2, I considered *fuzzy lies*, *i.e.* lies that involve *beliefs* about graded truth values. Utterances like (2\*), similarly, are insincere *statements* that express *beliefs* about graded truth values. This analogy suggests that, with respect to *fuzzy lies,* we have to consider two graded layers of insincerity: the layer of the speaker’s beliefs and the layer of the beliefs expressed by his statements. For instance, Bruce can tell a *fuzzy* *lie* either by ‘plainly’ stating (2) while believing that (2) is partly false (e.g. 0.3-true) or by stating (2\*) while believing that (2) is utterly false (that is, believing that (2\*) is partly false).

#### 4.5.2. Two directions of belief misrepresentation

A similar, but more complex discourse applies to the *degrees of belief* expressed by assertions. Assertions that express graded beliefs are generally overlooked in the literature on lying. This is because, in standard cases, statements express a flat-out belief in the truth of the proposition, rather than a graded belief. For instance, (3) expresses a flat-out belief in the asserted proposition:

* 1. Iraq has weapons of mass destruction

Not all statements, however, are as simple as (3), for some express graded beliefs. For instance, (3a) indicates that the speaker believes that (3) is probably true, and (3b) expresses uncertainty in the truth of the proposition:

(3a) (0.75/p) *Probably* Iraq has weapons of mass destruction

(3b) (0.5/p) *Maybe* Iraq has weapons of mass destruction

Few authors have raised the question of how assertions that express graded beliefs are to be analysed within a theory of lying. Meibauer (2014: 225) suggests that there are three kinds of graded insincere assertions that may qualify as lies: those “(i) expressing certainty when [you] are uncertain, those (ii) expressing uncertainty when [you] are certain, and those (iii) expressing certainty or uncertainty to a higher degree than being adequate with respect with [your] knowledge base”. Since the third case seems to include the previous two, to simplify this taxonomy I will simply distinguish between two ‘directions’ in misrepresenting your degree of belief: namely, pretending to have a *higher degree of belief* or a *lower degree of belief* than the one you have (cf. Falkenberg 1988:93).

A first, tempting idea is to assume that these two directions are equivalent. This would mean that, from the point of view of the analysis of lying, “pretending to be more certain than you are” is *as insincere as* “pretending to be less certain than you are”. A reason to make this assumption is that the ‘discrepancy’ between your state of mind and the state of mind expressed by the statement is the same in both cases. However, at a closer look this assumption reveals it to be naïve, as the first case (*overstating*) is often perceived as being more insincere, or more misleading, than the second (*understating*). To see this, consider the two utterances:

(3c) (1/p) *Certainly* Iraq has weapons of mass destruction

(3d) (0.5/p) *Perhaps* Iraq has weapons of mass destruction

Imagine that in both cases George’s mental state is in between certainty and uncertainty, so that he believes:

(0.75/p) [Probably Iraq has weapon of mass destruction]

According to the ‘naïve’ view, (3c) and (3d) are equivalent scenarios, because the discrepancy between BΨ and BΛ is the same (0.25). These scenarios differ only in the direction of misrepresentation: (3c) represents the speaker as having a higher degree of belief than he has, while (3d) as having a lower degree of belief. Interestingly, however, it is natural to assess (3c) as more insincere than (3d). The reason is that we tend to judge (3d) as a prudent statement, that cooperatively avoids saying more than the speaker knows, while (3c) is perceived a misleading overstatement, that the speaker has not sufficient knowledge to assert. In other words, *ceteris paribus*, understating your degree of belief is generally seen as a cooperative linguistic practice, while overstating it is generally regarded as uncooperative.

In line with this intuition, Falkenberg (1988: 94, 1990) proposes to distinguish between ‘hard lies’ (*overstatements*, like (3c)) and ‘soft lies’ (*understatements*, like (3d)). However, this taxonomy is misleading in two respects. First, not all overstatements and understatements are lies: if the CIC is a condition for lying, only statements displaying a certain level of discrepancy between BΨ and BΛ can be lies. Second, it is not clear whether an *overstatement* (hard lie) is necessarily more of a lie than an *understatement* (soft lie): the next section will show that the direction of misrepresentation is just one of the parameters of intensity that must be considered, another one being the magnitude of the discrepancy between BΨ and BΛ.

#### 4.5.3. Epistemic modals and degrees of commitment

The most prominent linguistic devices used to mitigate or reinforce the degree of belief expressed by an assertion (expression like ‘certainly’, ‘probably’, ‘perhaps’) are called *epistemic* *modals*. This section will analyse how they alter the degree of belief expressed by the assertion, and clarify why we generally assess *understatements* as more sincere (or more honest) than *overstatements*.

On a pragmatic level, epistemic modals both “indicate the *speaker’s confidence* or lack of confidence in the truth of the proposition expressed” and “*qualify [his] commitment* to the truth of the proposition expressed in [his] utterance” (Coates 1987:112, italic is mine). In other words, they act on two components of the assertion, altering both (1) the psychological state expressed by the speaker (the degree of belief), and (2) his degree of commitment to the truth of the proposition (the illocutionary strength[[76]](#footnote-76)) (cf. Sbisà & Labinaz 2014:52, Lyons 1977: 793-809; Holmes 1984: 349).

These two functions are distinct in nature, but entangled: if a speaker S mitigates (or reinforces) the degree of belief conveyed by his assertion, then S automatically mitigates (or reinforces) the illocutionary force of his assertion (that is, his degree of commitment to the truth of the proposition). For instance, if you state (4b) instead of plainly stating (4), you both mitigate the degree of belief expressed ((4b) expresses uncertainty in (4)) and lower the degree of your commitment to the truth of the asserted proposition (you are committed to the truth of (4) to a much lower degree if you utter (4b))[[77]](#footnote-77).

(4) Plato will quit smoking tomorrow

(4b) *Perhaps* Plato will quit smoking tomorrow

The role that epistemic modals play in reinforcing/weakening the illocutionary force of assertions explains why *understatements* are perceived as more honest than *overstatements*. *Ceteris paribus* (given the same degree of insincerity, like in (3c)-(3d)) a reinforced assertion has a stronger illocutionary force than a mitigated assertion, so that the speaker has a stronger commitment to its truth. And if the commitment to sincerity is stronger in reinforced statements, then violating that commitment is more serious in those statements than in mitigated ones.

Variations in illocutionary force induced by epistemic modals can affect whether the speaker is asserting the proposition or not – and hence whether he is lying, because lying requires asserting. This is because epistemic modals can downgrade the degree of illocutionary force of a declarative sentence to such an extent that it longer counts as an assertion, but rather as a supposition or a hypothesis (Sbisà & Labinaz 2014:52-3). For instance, (4b) is a supposition rather than an assertion: its insincere utterance does not amount to lying, while insincerely uttering its unmitigated version (4) does. Carson (2010: 33,38) shares this intuition: “there are weaker and stronger ways of warranting the truth of a statement. To count as a lie, a statement must be warranted to a certain minimum degree”.

This is even more evident in other speech acts. For instance, if Matteo utters (5b) instead of (5), it is clear that he has not promised that he will buy you an elephant (he is merely suggesting it), while he would be promising it if he uttered (5). It seems that an insincere utterance of the first amounts to lying, while this is not true for the second[[78]](#footnote-78).

(5) Tomorrow I will buy you an elephant

(5b) *Perhaps* tomorrow I will buy you an elephant

This theoretical framework allows us to correctly analyse lying involving assertion that express graded beliefs (BΛ-graded lies, and complex cases where both BΛ and BΨ are graded). The problem with these assertions is that they cannot be dealt with simply by appealing to condition CIC. Without an account of how epistemic modals modify illocutionary force, CIC alone cannot account for the differences determined by the direction of misrepresentation (overstatements vs understatements). This difficulty dissipates once it is understood that epistemic modals influence not only whether the sincerity condition is satisfied (by altering the *degree of belief* expressed), but also whether the assertion condition is satisfied (by altering the speaker’s *degree of commitment*).

To sum up, there are three entwined scalar parameters that we have to consider when we analyse the graded components of the belief expressed by an utterance: the *graded truth value* expressed by the utterance, the *graded belief* itcommunicates (these two parameters determine the degree of insincerity), and the *degree of assertoric force* of the utterance, or degree of commitment (this parameter determine if the speaker is asserting, and how strongly) that the speaker undertakes. For an utterance to count as a lie, both a certain degree of insincerity *and* a certain degree of commitment must obtain.

### 4.6 Further graded components

So far, I have considered aspects of intensity that are properly linguistic. Several paralinguistic devices can intervene to modulate the illocutionary force of an assertion. A significant factor is *prosody*, and specifically *intonation*: Gussenhoven (2002) remarks that variations in volume and tonal height are a powerful device to communicate the speakers’ confidence in the truth of his assertion. Among other influential factors that can be used to influence the strength of an utterance are pauses, rhythm of speech, repetitions and proxemic signals.

Context is also an important factor to be considered. Bazzanella (2009, 78) points out that, among the parameters affecting the degree of insincerity of a lie, also “aspects of the global context [...], of the cotext [...] and of the local context” are of central importance. For instance, an utterance like (1) can express a different degree of certainty whether (1) is uttered in (context A) a restaurant, where a person addresses it to a dining companion that is known not to like cheese, or (context B) in a hospital, where a doctor addresses it to a person that he knows to have a deadly allergy to cheese. In both cases, given that (1) is insincere, (1) is a lie, but it is more intense in the second case, since the commitment to the truth of the assertion is stronger in context A than in context B.

* 1. This meal contains no cheese

The very mitigating and reinforcing devices considered so far are influenced by contexts – in most cases, their very meaning is determined by contextual factors (Kratzer 1981). For instance, the epistemic modal ‘definitely’ in (2) expresses a different degree of certainty and commitment depending on whether (2) is uttered in context A or in context B.

* 1. Definitely, this meal contains no cheese

Linguistic, paralinguistic and contextual elements determine a complex interplay of factors that influences the degree of illocutionary force of an assertion, and thus the degree of intensity of a lie. However, it seems to me that only linguistic devices (like epistemic modals) are strong enough to affect the status of an utterance as an assertion (and therefore as a lie). Paralinguistic and contextual elements can reinforce or mitigate an assertion, but they alone cannot determine if an utterance counts as an assertion or not.

In table 1, I summarise my taxonomy of the layers of gradedness involved in lying. This taxonomy is not meant to be exhaustive, but it succeeds in underlining a fact that has been ignored by the philosophical literature on lying for a long time: the multi-layered gradedness of lying.

|  |  |
| --- | --- |
| **Graded of *beliefs*** | Beliefs about graded truth values |
| Graded beliefs about truth values |
| **Gradedness of *statements*** | Stating “beliefs about graded truth values” |
| Stating “graded beliefs about truth values” |
| Modifying illocutionary strength |
| **Gradedness of *paralinguistic components*** | Modifying illocutionary strength |
| **Gradedness of *contexts*** | Modifying illocutionary strength |

**Table 1**: A taxonomy of the dimensions of gradedness involved in lying.

In conclusion, the gradedness of lying results from the interaction several parameters: on the side of beliefs (BΨ), graded beliefs and beliefs about graded truth values; on the side of statements expressing such beliefs (BΛ), the numerous ways to convey information about them in statements. In the literature on lying, scholars tend to completely ignore these graded features when they assume the dichotomic view as a starting point – a view that I proved to be inconsistent with our intuitions.

I have provided several reasons to believe that the traditional insincerity condition is wrong, because it rules out *fuzzy lies* and *graded-belief lies*. Moreover, I have shown that the traditional insincerity condition yields a wrong descriptionof lying, as it blinds us to its graded nature. My proposed definition corrects this picture and allows for *fuzzy lies* and *graded-belief lies* to be counted as genuine lies*,* and acknowledges the existence of many degrees of insincerity in lying. I have also shown that the very modifiers that affect the degree of belief expressed by an assertion modify its *illocutionary* force, affecting whether a given utterance counts as an assertion.

My main aim in this section has been to show how the insincerity condition for lying need to be modified in order to correctly deal with cases involving graded insincerity. It should be kept in mind, however, that also other components of lying have graded features. The picture is complex: as Bazzanella (2009, 78) points out, “the different degrees of intensity in lying result from the complex interplay of various layers and parameters”. For instance, some authors (Chisholm & Feehan 1977; Fallis 2011; Staffel 2011; Marsili 2017) contend that the intention to deceive, as well as the effects of deception, can be graded. Several pragmatic parameters determinant for lying are also graded, such as relevance (Sperber & Wilson 2002; Van der Henst et al. 2002), felicity, and the relations between the interactants (like social relations, and respective trust) (Bazzanella 2009).

Finally, moral evaluations of lying can be graded – interestingly, in a way that can parallel the graded components of lying just identified. Intuitively, some lies are worse than others: lying to save a life is better than lying to get away with murder. Many factors will intervene in determining the reprehensibility of a lie: two obvious candidates will be the *effects* of the lie and the *intentions* of liar (quite obviously, good effects and intentions are preferable to bad ones). But especially from a *deontologist* perspective (Augustine *DM*, Aquinas *ST*, Kant *GMM, SRTL*, Newman 1880, Geach 1977, Pruss 1999, Tollefsen 2014, cf. also Isenberg 1964), a great deal of what is wrong with lying is that lying violates a moral norm of sincerity. In pointing out that such norm can be violated to a greater or lower extent, and in developing a formal model for measuring and describing the extent of such violations, my proposed account represents also a valuable tool for assessing the (im)morality of lying.

## 5. Conclusions – A general account of insincerity

So far, in my discussion of the graded nature of insincerity I have ignored attitudes other than beliefs. In this final section, I will show how my discussion generalises to these other attitudes. As a starting point, let us reconsider the general speech-act theoretic account of insincerity presented in 2.2:

INS: The performance of an illocutionary act F(p) that expresses the psychological state Ψ(p) is insincere IFF in uttering F(p), S is not in Ψ(p)

Applied to beliefs, INS gives us the following belief insincerity condition (BIC):

BIC: S asserts that p insincerely only if S does not believe that p

I have already discussed condition BIC (presented as Carson’s weak insincerity condition, or WIC) in detail, showing that BIC is able to deal with cases of graded insincerity, like *fuzzy lies* and *graded* lies. However, I rejected BIC as a necessary condition for lying, on the ground that it incorrectly counts *bullshitting* (saying something you neither believe to be true nor false) as lying. Since INS implies that the insincerity condition for assertion is BIC, INS is not well suited as a general account for the insincerity conditions for lying.

On the other hand, in the IV.1 I mentioned that I am concerned with two related notions of insincerity: an account of the *insincerity condition in the definition of lying*, and an analysis of *insincerity simpliciter*. We have already seen that these two notions can come apart: for instance, speech acts that do not entail assertions (e.g. *requests, orders*) can be insincere but cannot be lies. INS is not viable as an account of the *insincerity conditions for lying*, but it seems an appropriate analysis of insincerity *simpliciter*. This is because bullshitting falls short of lying, but is arguably a form of insincere speech. If one shares this intuition, then bullshit is not a counterexample to BIC or INS understood as characterising insincerity *simpliciter,* rather than as a necessary condition for lying. If one, by contrast, intuits that bullshitting amounts to neither lying nor insincerely speaking, the general account of the insincerity condition for lying that I am about to delineate will coincide with their desired account of insincerity *simpliciter*, as both exclude bullshitting from the definition.

How can we narrow INS so that it excludes bullshitting but still generalises to speech acts other than assertion? In the previous section I have argued that the comparative insincerity condition CIC is preferable to BIC as a necessary condition for lying:

CIC\*: S is more confident in ¬p than he is confident in p

The desired refinement of INS must entail CIC\* rather than BIC, and generalise to speech acts that express attitudes other than beliefs. To be sure, at least when it comes to insincerity conditions for lying, we do not need to show that INS generalises to every possible attitude. In chapter III, I have shown that only commissive and assertive illocutionary acts can entail an assertion and thus count as lies (III.4). Since these families of speech acts only express either *beliefs* or *intentions* (Searle 1976), a general account of the insincerity conditions for lying only needs to generalise to intentions and beliefs. In other words, our minimal desideratum is an extension of the graded insincerity conditions for beliefs (CIC\*) to intentions.

It seems that we can reformulate the intention insincerity condition (IIC) in a way that parallels CIC\*, namely in a way that involves a comparison between a psychological state and its opposite:

IIC: S does not intend to p

IIC\*: S intends to not-p more than S intends to p

Admittedly, the IIC\* is a bit odd-sounding. One reason is that intentions do not seem to come in degrees in the same way belief do (but cf. Holton 2008). Intuitively, intending is an ‘on/off predicate’: either one intends to eat an apple, or one does not intend to eat an apple – it is just not clear which intermediate mental state could exist between the two. If this is right, it is not clear how allowing for intermediate cases between an intention and its opposite can introduce a meaningful refinement of IIC.

Perhaps IIC\* can be interpreted as involving a comparison between graded truth-values, rather than graded intentions. In other words, the point of IIC\* would be to capture promises that express insincere intentions akin to *fuzzy lies*, rather than *graded-belief* lies. Suppose for instance that I promise to my girlfriend:

* + 1. I promise (1\*) [that I will get fit]

In promising (1), I express an intention with content (1\*), namely an intention to get fit. It seems that this intention can be insincere to different extents: I can intend to definitely get fit (in which case (1\*) has value of 1), intend to definitely not get fit (in which case (1\*) has a value of 0), and I can have a number of intermediate intentions, depending on the graded truth-value assigned to (1\*) – for instance, intending to get *somewhat* fit (in which case (1\*) would have a value of, say, 0.3), get *quite* fit, etc. Unlike IIC, IIC\* allows for these intermediate states. On top of this, it individuates a plausible boundary between these intermediate states, for the same reason that CIC\* individuates the right boundary for beliefs – namely, it seems that the limit between insincerity and sincerity does not lie at any arbitrary point close to the extremes, but rather in between them.

This is good news, as it shows that the CIC\* can be neatly extended to attitudes other than beliefs. We can thus derive the following general insincerity condition for lying, from which both IIC\* and CIC\* can be derived:

**Graded insincerity condition for illocutionary acts**

INS-L: The performance of an illocutionary act F(p) is insincere IFF in uttering F(p), S is in Ψ(¬p) more than S is in Ψ(p)

INS-L is designed to deal with cases involving intentions or beliefs, but it also generalises to other attitudes, such as desires. If one has the intuition that *bullshitting* is not a form of insincerity, INS-L will also offer a plausible characterisation of insincerity *simpliciter*. More importantly, INS-L offers the general background that is needed to offer a general definition of lying that applies to speech acts other than assertions and attitudes other than beliefs. Such a definition, derived by integrating INS-2 into the definition of lying developed in chapter III, reads as follows:

**Speech act theoretic definition of lying**

*In successfully uttering an illocutionary act with content p that expresses an attitude* *Ψ(p) and entails an assertion with content p, S lies to A about p iff:*

1. *S thereby asserts that p (i.e. conditions (a), (b), and (c) from III.3.1 obtain)*
2. *Either S is more confident in p than S is confident in ¬p, or S is in Ψ(¬p) more than S is in Ψ(p), or both*

In the case in which the speaker is asserting directly, this definition will reduce to the following:

**Definition of lying by asserting**

*In successfully uttering an illocutionary act with content p*, *S lies to A about p iff:*

1. *S thereby asserts that p (i.e. (a), (b), and (c) obtain)*
2. *S is more confident in ¬p than S is confident in p*

This completes this dissertation’s reflection on the definition of lying. My proposed definition is perhaps complex and thus less elegant than competing ones, but it gives better predictions than the alternative accounts, and finds independent theoretical support from a general theory of illocutionary acts. Unlike deceptionist accounts, it is not subject to counterexamples to the intention to deceive condition. Unlike assertion-based accounts, it deals correctly with lies involving explicit performatives. And unlike any other account, it is able to capture to speech acts other than assertions and attitudes other than beliefs, and to include both *graded-belief lies* and *fuzzy lies*.

Thus far, I have provided a characterisation of what it means to lie and to be insincere. Avoiding lying and insincerity, however, is not all that we require from our interlocutors. In the next chapters, I will deal with another expectation that is fundamental for our communicative interactions, namely expectations about the epistemic standpoint of our interlocutors. When someone asserts something, we generally expect such claim to be backed up by some reasons, rather than mere belief – in other words, all things the same, a ‘gut feeling’ that a proposition is true is not enough for it to be assertable. However, it is not clear exactly which kind of epistemic standpoint is required for an assertion to be permissible *qua assertion*. The next chapter will deal the ‘epistemic norm of assertion’ hypothesis, namely the idea that assertions are subject to a norm of the form: “assert that *p* only if *p* has C”, where C indicates the unique epistemic property that a proposition must have for it to be assertable.

**Part B - The norms of assertion**

# B. The norms of assertion

Quisquis autem hoc enuntiat quod vel creditum animo, vel opinatum tenet, etiamsi falsum sit, non mentitur. Hoc enim debet enuntiationis suae fidei, ut illud per eam proferat, quod animo tenet, et sic habet ut profert. Nec ideo tamen sine vitio est, quamvis non mentiatur, si aut non credenda credit, aut quod ignorat nosse se putat, etiamsi verum sit: incognitum enim habet pro cognito.

Now whoever utters that which he holds in his mind either as belief or as opinion, even though it be false, he lies not. For this he owes to the faith of his utterance, that he thereby produce that which he holds in his mind, and has in that way in which he produces it. Not that he is without fault, although he lie not, if either he believes what he ought not to believe, or thinks he knows what he knows not, even though it should be true: for he accounts an unknown thing for a known.

Augustine, De Mendacio, III.3

In the previous chapters, I have provided a detailed analysis of a common communicative vice: lying, understood in terms of asserting something that you believe to be false. In this chapter, I will discuss another kind of communicative vice: that of asserting something in absence of the appropriate warrant or epistemic support.

Once again, Augustine’s *De Mendacio* offers an interesting starting point to introduce the debate: in the passage quoted above, we can find what is arguably the first philosophical discussion of assertions that are sincere, but nonetheless unwarranted. Here Augustine is discussing the distinction between *inadvertently saying something false* and *lying*. We have already seen that this distinction is intuitively correct: lying and being mistaken about what you say are two different concepts. Importantly, here Augustine finds it necessary to specify that while *inadvertently saying something false* falls short of lying, also this kind of assertion involves a *vitium ­­*­- a fault, violation or vice.

What kind of *vitium* or faultexactly? There are many faults that we might identify in such assertions: lack of epistemic support, or more simply lack of accuracy, or of correspondence with reality. To make the discussion more concrete, let us consider an example. Suppose that Claudia and Rachel are having a conversation about their common friend Jacques, and Rachel asks Claudia whether Jacques is a good cook. Claudia has never tasted Jacques’ cuisine, nor has she got any second-hand information about Jacques’ abilities in the kitchen. However, Claudia knows that Jacques is French, and she is under the impression that French people are generally good cooks. On this basis, she replies:

1. Sure, Jacques is a good cook

Suppose that Jacques is not, as a matter of fact, a good cook: he is utterly terrible in the kitchen. In this case Claudia is not lying; nonetheless, her assertion is not “without fault”, as Augustine would say. But which kind of fault is exactly involved in Claudia’s assertion that (1)?

A contemporary approach to explaining what is wrong with (1) is to point out that the speaker (in this case, Claudia) failed to meet some relevant conversational norm, and consequently some conversational expectation. Arguably, assertors should have some epistemic ground to support what they say: there seem to be an implicit norm that dictates that they should meet some minimum epistemic standard before they assert something. In our example, (1) is false, and Claudia is not justified to believe that (1) is true. One way to explain what is wrong with (1), then, is to say that in uttering it Claudia violates a putative conversational norm – for instance, a norm requiring that you only utter true (or known, or reasonably believed) propositions. But Claudia’s assertion could also violate some other conversational expectation. In asking whether Jacques is a good cook, Rachel is expecting Claudia to base her assertion on some appropriate grounds, not on a wild guess; she expects her to assert a true proposition, not a false one. At least on a first intuitive level, it seems a reasonable hypothesis that we perceive wild guesses like (1) to be faulty because they violate some relevant conversational norm, or at least the expectations that might be reasonably be held by the participants in the conversation.

In the last twenty years, the hypothesis that the faultiness of assertions like (1) might be explained in terms of the violation of a putative ‘norm of assertion’ has gained centre stage in philosophy of language and epistemology, sparking a lively debate around the question of which norm regulate assertions. Timothy Williamson’s “Assertion” (1996, revised in 2000) is the perhaps the first paper to explicitly and systematically address this question, and definitely the one responsible for initiating the contemporary debate over this issue. Williamson opens his paper by putting forward a simple hypothesis (Williamson 2000:241):

Williamson’s Hypothesis

*What are the rules of assertion? An attractively simple suggestion is this. There is just one [constitutive]*[[79]](#footnote-79) *rule. Where C is a property of propositions, the rule says:*

*(The C-rule)* • *One must: assert p only if p has C.*

According to this hypothesis, there is only one rule, the C-rule, to which all and only assertions are subject – a rule that tells you which propositions you can properly assert and which ones you cannot. This rule requires you to assert only propositions that have the unspecified property ‘C’.

Williamson clarifies that the C-rule (1) is *unique* to assertion: *only* assertion is regulated *only* by this rule. Since only assertions are regulated only by this rule, the C-rule also (2) it *individuates* assertion, defining it as the only speech act that is only subject to the rule. Furthermore, the C-rule (3) is *constitutive* of assertion: were assertion regulated by a different rule, it would be a different speech act. Furthermore, and relatedly, the C-rule is (4) a norm to which assertors are subject *qua assertors*. There are many normative constrains that can contribute in making an assertion *overall* wrong: an assertion can inappropriate because it is impolite, immoral, rude or irrelevant. While these are all good reasons not to proffer an assertion, they are not infractions that of a norm to which *only* assertions are subjects: commands, questions (and perhaps even some non-communicative behaviours) can be impolite, immoral, rude or irrelevant too. The norm that Williamson is looking for, by contrast, is a norm to which *only* assertions are subject, and to which assertors are subject *in virtue* of their being asserting something.

Williamson’s hypothesis is indeed “attractively simple”: if correct, we can both define what assertion is and specify under which condition its performance is appropriate, simply by identifying one property, C. This raises the question that animates the debate: what kind of property is C? Williamson’s answer is that C is the property of *being known by the assertor*. In other words, assertion is governed by the norm that one should not assert what one does not know to be true:

Williamson’s Answer: the Knowledge-Rule

KR: “You must: assert that p only if you know that p”.

A number of philosophers have found KR a convincing answer (e.g. DeRose 2002; Hawthorne 2004; Benton 2014). However, the debate is far from settled, and Williamson’s position has elicited a number of critical responses. One could roughly divide these critical reactions into two categories.

The first category comprises those who reject or challenge Williamson’s hypothesis, either in part or as a whole. Some philosophers (Brown 2008, Carter 2014; 2017, Carter & Gordon 2011, Gerken 2014, McKenna 2015) reject the assumption that there is *only one* norm of assertion. Others deny that assertion is regulated by a norm (Sosa 2009, cf. also Rescorla 2007; 2009), or that it is *constituted* by it (Hindriks 2007, Maitra 2011, Pagin 2011, McCammon 2014:137-9). Finally, some have gone so far as to claim that there is no such thing as an ‘assertion-game’ to which the putative rule applies (Cappelen 2011, Johnson 2017).

In the next chapter, I address criticisms of this kind. More specifically, I object to the claim that the C-rule is ‘constitutive’ of assertion. After reviewing several difficulties for this hypothesis, I show that abandoning the idea that the norm of assertion is constitutive also puts strain on the assumption that the norm is *unique* to assertion, and consequently on the assumption that assertion can be defined as the only speech act subject to this rule.

The second category of criticisms comprises those coming from scholars who accept Williamson’s hypothesis, but refuse Williamson’s answer in favour of a different one, *i.e.* an alternative account of what property C is. For instance, some maintain that a warranted assertion requires instead the *truth* of the proposition (Weiner 2005, Whiting 2012), or some relevant *reason to believe it* (Lackey 2007, Kvanvig 2009).

In chapter VI, I consider criticisms of this kind. More specifically, I address the disagreement between factive and non-factive accounts, *i.e.* accounts that do (like Williamson’s knowledge-rule) or do not require that the proposition is true. After presenting some objections to factive accounts, I present my own non-factive proposal. On this view, false assertions are faulty because they fail to meet the success-condition for asserting something (the purported aim of assertion), rather than a permissibility condition (the norm regulating assertion).

# V. A ‘constitutive’ norm?

According to Williamson’s influential hypothesis, assertion is subject to a single constitutive norm, taking the form: “One must: assert that *p* only if *p* has C”. Some philosophers, however, have challenged the idea that a rule of this kind can be considered constitutive of assertion. In this chapter, I present my own version of this criticism.

In section 1 and 2, I introduce two notions of ‘constitutive rule’: the orthodox notion and Williamson’s notion. I show that Williamson’s notion departs significantly from the orthodox one, and I introduce a difficulty: problematically, not all authors writing on the norm of assertion adopt Williamson’s understanding of the notion of ‘constitutive’. Some endorse the orthodox interpretation; other Williamson’s; others hybrid notions. I then review these different approaches, attempting to determine which one is better suited to make sense of the claim that assertion is governed by a constitutive norm. In section 3, I explore interpretations that are closer in spirit to an orthodox understanding of the notion, and I show that they are structurally incompatible with the norm of assertion hypothesis. In section 4, I explore interpretations that are closer in spirit to Williamson’s understanding of the notion. The problem of this views is that they do not allow for a meaningful distinction between constitutive and non-constitutive rules, so that claiming that the norm of assertion is a constitutive rule in this sense merely amounts to claiming that the norm of assertion is a rule. In section 5, I conclude that there is no such thing as a *constitutive* norm of assertion – at least not in a sense that can be meaningful for the current philosophical debate. As we abandon this assumption, others lose their appeal: namely, the assumption that assertion is governed by *only one* norm, and the assumption that assertion can be identified as the only speech act subject to this norm.

## 1. The orthodox account of constitutive norms

Williamson claims that the norm of assertion is a *constitutive* rule: “[My] paper aims to identify the constitutive rule(s) of assertion, conceived by analogy with the rules of a game. [...] Henceforth, ‘rule’ will mean constitutive rule” (1996:489-90). The notion of ‘constitutive rule’ is widely employed across a range of different philosophical subjects, “as diverse as philosophy of language, philosophy of law, and most recently artificial intelligence” (Hindriks 2009). Authors employing this notion rarely call into question the *orthodox account*: the account of norms that was developed by Searle (1964, 1969, 1995), as a refinement of Rawls’ account (1955) – adopted, amongst others, by Lewis (1979). In this section, I will cover the orthodox account; in the next one, I will move on to Williamson’s.

On the orthodox view, constitutive rules are defined in opposition to *regulative rules*. Regulative rules track our ordinary understanding of what a rule is: an imperative that serves as a “guide, or as a maxim, or as a generalization from experience” (Rawls 1956:24). Imperatives of this kind can be violated (it is possible to disobey the rule), but transgressors are *prima facie susceptible* to criticism (or sanctions) for violating the rule. Given their directive function, regulative rules “characteristically have the form or can be comfortably paraphrased [as imperatives] of the form ‘Do X’ or ‘If Y do X’” (Searle 1969:33). For instance, “Do not chew with your mouth open”, or “if the traffic light turns red, vehicles must stop”.

Constitutive rules differ from regulative rules in several ways. A first distinctive aspect is that they are more like definitions than commands. This is because they do not set obligations; rather, they define what it is to engage in a particular institutional practice:

Regulative rules regulate antecedently or independently existing forms of behaviour [...]. But constitutive rules do not merely regulate, they create or define new forms of behaviour. The rules of football or chess, for example [...] create the very possibility of playing such games. (Searle 1969:33)

Constitutive rules can be phrased as ‘count-as’ locutions: “X counts as Y-ing in C” (1969:33). For instance, the constitutive rule of checkmate can be expressed as follows: “attacking the king in such a way that no move will leave it unattacked counts as checkmate in the game of chess”. The ‘definitional’ (rather than imperative) character of constitutive rules reflects their ‘performative’ function: they establish institutional practices by defining them (Searle 1995).

A consequence of the non-imperative character of constitutive rules is that while regulative rules can be disobeyed, constitutive rules cannot: “it is not easy to see how one could even violate the rule as to what constitutes checkmate in chess, or a [goal] in football” (Searle 1969: 41). If a practice is established by a constitutive rule, that practice can only be properly described as such if it follows the rule. Hence, by not following a constitutive rule, rather than violating the rule one ceases to be engaged in the activity subject to that rule: for instance, checkmating the opponent incorrectly simply amounts to a failure to checkmate.

To sum up, the orthodox account claims that: (1) constitutive rules establish the practices they regulate, while regulative rules regulate pre-existing practices; (2) given their nature, constitutive rules cannot be violated, while regulative rules can; (3) constitutive rules can be easily formulated as count-as locutions, while regulative rules generally take the form of imperatives. The orthodox account is summarised in the following table:

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **constitutive rules** | **#** | **regulative rules** |
| **C1** | Establish (define) the practices they regulate | **R1** | Regulate pre-existing practices |
| **C2** | Cannot be ‘properly’ violated | **R2** | Can be violated |
| **C3** | Typically take the form of ‘count-as’ locutions | **R3** | Typically take the form of imperatives |

Table 1: A summary of the differences between constitutive and regulative rules, according to the orthodox account

## 2. Williamson on constitutive norms

In presenting his account of constitutive rules, Williamson significantly departs from the orthodox understanding of constitutive rules: his analysis clashes with (C2) and (C3). Since this difference is rarely acknowledged in the literature, it is worth pointing out that Williamson himself fails to mention it: on the contrary, he presents the proposed interpretation as uncontroversial (“we have at least a crude conception of constitutive rules...” 2000: 239) and in line with the tradition.

Williamson refuses to commit himself to a particular definition of constitutive rules by stating that he will make “no attempt […] to define ‘rule’” (2000: 239). Nevertheless, he gives a necessary but not sufficient condition for being a constitutive rule: “A rule will count as constitutive of an act only if it is essential to that act: necessarily, the rule governs every performance of the act” (2000:239).

A core tenet of this view is that, if a rule is constitutive of an act, it is *essential to* it. Being ‘essential to’ the act means that the rule *necessarily* governs that act (rather than contingently, or as a result of a mere convention): assertion could not have been governed by another norm, and it could not exist if such a rule was not in place (cf. Pagin 2015: §6.2). In requiring that the rule *necessarily* governs the act, Williamson confers a modal flavour to (C1) that is absent in the orthodox conception: I will use (C1\*) to indicate this difference, which I shall revisit in section 5.

Williamson departs from the orthodoxy in maintaining, against (C3), that constitutive rules typically take an *imperative* form: in other words, he takes them to possess the regulative property (R3) rather than the constitutive property (C3) (cf. Maitra 2011: 280-4, Hindriks 2007: 396). He argues that constitutive rules characteristically put an agent under an obligation: “‘must’ expresses the kind of obligation characteristic of constitutive rules”. He phrases the C-rule in an imperative form: “In the imperative, assert *p* only if *p* has C” (2000: 241).

Williamson also departs from (C2), claiming that constitutive rules “do not lay down necessary conditions for performing the constituted act”. Consequently, satisfying the constitutive rule of assertion is not necessary for asserting: “when one breaks a rule of assertion, one does not thereby fail to make an assertion” (2000: 240). More generally, one can φ without satisfying the constitutive rule of φ-ing. So failing to satisfy the rule involves being liable to criticism, rather than not asserting. By contrast, on the orthodox view an infraction of a constitutive rule amounts to not φ-ing, rather than φ-ing inappropriately. Again, the proposed view is rather in line with the corresponding regulative property (R2).

An interpretative problem emerges here. On the one hand, Williamson attributes characteristically regulative properties (R2, R3) to the C-rule instead of the corresponding constitutive properties (C2, C3), so that the rule is not taken to be constitutive in the traditional sense (as pointed out by Hindriks 2007:396, McCammon 2014:137-9, cf. also Maitra 2011). On the other hand, several authors take the C-rule to stand out as the constitutive norm of assertion, in the customary sense (e.g. García-Carpintero 2004:143, Rescorla 2007:253, 2009:99, Turri 2013:281, Sbisà 2016).

This is not merely a terminological issue: the two notions are incompatible. Since different authors adopt different notions of ‘constitutive norm’, we are dealing with a crucial (and often unacknowledged) disagreement regarding the nature of the norm of assertion. This disagreement threatens to undermine the foundations of the debate: if there is no consensus about what ‘constitutive’ means, different authors can (and do) mean different things when they refer to the ‘constitutive’ norm of assertion, effectively talking past each other.

Solving the disagreement is not as simple as claiming that authors endorsing the orthodox reading are mistaken, and that Williamson simply interprets the notion in a novel way, namely as possessing properties (C1\*, R2, R3). In the course of this chapter, I will show that this solution is also problematic, as it renders the claim that the norm of assertion is ‘constitutive’ of little or no significance. Before entering the details of the pros and cons of each account, however, I want to introduce more in detail both proposed interpretations.

In the next section, I will try to interpret the C-rule as an orthodoxly constitutive rule that possesses some regulative features (R2, R3); in the subsequent one, as a regulative rule that possesses the constitutive feature (C1\*) instead of (R1). Both attempts will prove problematic and incompatible with the desiderata underlying Williamson’s hypothesis. I will conclude that the problem rather lies in the hypothesis: while assertion is arguably subject to a norm taking the form of the C-rule, this norm is not constitutive of assertion in any sense that is meaningful for the debate.

## 3. The C-Rule as a Constitutive Rule

### 3.1. Phrasing the C-rule as orthodoxly constitutive

A first strategy to make sense of the claim that the norm is ‘constitutive’ is to follow the authors who interpret the C-rule as a constitutive rule, on the orthodox understanding (García-Carpintero 2004:143, Rescorla 2007:253, 2009:99, Turri 2013:281, Sbisà 2016, cf. also Maitra 2011: 283-4, Goldberg 2015: §1). Can we show that there is a sense in which this interpretation is compatible with Williamson’s Hypothesis? The first difficulty is that, according to the Hypothesis, the C-rule rather takes a regulative form:

(A1) One must: assert that *p* only if *p* has C

To treat (A1) as a constitutive norm, perhaps we could rewrite it as follows:

(A1\*) One asserts that *p* only if *p* has C

(A1\*) satisfies (C1), (C2), and (C3), so that is a constitutive reading of (A1). However, (A1\*) is clearly not a plausible rephrasing, at least not in a sense relevant for the debate. In fact, most authors want to claim that property C is a property like K: ‘being *known* by the speaker’, or B: ‘being *believed* by the speaker’. But if C=K, or C=B, then saying what you do not know, or what you do not believe, is not asserting – a possibility explicitly denied by Williamson (2002: 240). Since (A1\*) is in plain contradiction with the hypothesis, according to which an assertion in violation of the norm is still an assertion, I will turn to another constitutive reading that avoids this difficulty.

### 3.2. An alternative constitutive reading

To turn (A1) into a constitutive norm, one needs a rephrasing that satisfies (C1), (C2) and (C3) and that acknowledges the existence of assertions without property C.

Such a phrasing could take the following form:

(A2) One asserts that *p* iff in asserting *p*, one is subject to the obligation that *p* must have C[[80]](#footnote-80)

This phrasing seems to satisfy our desiderata, but it clearly contains a circular element, as assertion figures on both sides of the biconditional. Addressing this worry, however, can be as simple as setting the relevant constraint on a less specified activity – in our case, on sayings rather than assertions. Assertions can be thus defined as the only saying (or speech act) that is subject to the unique norm that what is saidmust have property C. The requisite that *p* be said is meant to avoid the counterintuitive prediction that presuppositions and implicatures can be assertions (cf. Alston 2007:24-5, Pagin 2015:§2.2):

(A3) One asserts that *p* iff in saying *p*, one is subject to the obligation that *p* must have C

Now, (A3) differs significantly from the formulation (A1) originally found in Williamson, so that it represents a substantive amendment of his view. And it is far from clear that any author writing on assertion would endorse this particular interpretation of Williamson’s Hypothesis. Nonetheless, there are strong reasons to prefer this reading, as it appears to meet all the desiderata of the orthodox conception. Consistently with (C1), it *defines* the practice of assertion by declaring what it is to engage in that practice: it is to be *subject to* the obligation that what is said must have C. Consistently with (C2), it cannot be disobeyed: if one is not subject to the obligation that *p* has C, one is simply not asserting. And, consistently with (C3), it can be phrased as a count-as locution[[81]](#footnote-81). Moreover, it allows for the existence of assertions that do not have property C, thus avoiding the problem of (A1\*).

### 3.3 Treating two rules as one

Interpretation (A3) is appealing because it apparently fits both the orthodox view and Williamson’s. However, there is patently something wrong with this apparent reconciliation: the orthodox interpretation of ‘constitutive norm’ requires (C2) and (C3) to be satisfied, and this is inconsistent with Williamson’s understanding of the C-rule, that *rejects* (C2) and (C3) in favour of (R2) and (R3). No consistent conception of a rule can meet both requirements, as requirements (C2) and (R2), and (C3) and (R3), are mutually exclusive: (R2) allows for violations of the norm that are still assertions, while (C2) does not; (R3) takes the rule to be akin to an imperative, while (C3) does not.

Here is a possible reply. Even if (A3) as a whole cannot be violated in asserting, there is a sense in which (A3) expresses an obligation that can be violated: (A3) states that a proposition should be asserted only if it has property C. It is then possible to state a proposition that does not possess property C, while still asserting. This shows that (A3) possesses not only constitutive features (C1-3) as shown in the previous section, but also regulative features (R2-3) – thus meeting the requirements of both the orthodox and the Williamsonian notion.

Though appealing, this reply rests on a mistake (perhaps the most common mistake in interpreting Williamson’s Hypothesis): on this reading, two rules are treated as one. This is because (A3) is a rule about another rule: it is a constitutive rule that states that if you assert, you are subject to another rule, a regulative rule, (A1). To show this, (A3) can be paraphrased as:

(A3\*) One asserts that *p* iff in saying *p,* one is subject to rule (A1)

Where (A1) is:

(A1) One must: assert that *p* only if *p* has C

It is now apparent that (A3) does not possess both regulative and constitutive properties. While (A1) can be violated, (A3) cannot be violated; (A1) is an imperative, whereas (A3) is a (non-imperative) definition. To claim that (A3) possesses both regulative and constitutive features is to conflate (A3)’s properties and (A1)’s properties. This is a mistake, as (A3) and (A1) are different rules with different properties. (A3) is a definition of assertion and cannot be violated, while (A1) is an imperative and can be violated; the first is a constitutive rule in the traditional sense, the second is not[[82]](#footnote-82).

The confusion between these two rules is particularly significant, as it can explain why different authors endorse different interpretations of the claim that the norm is constitutive – (A1) allows for a regulative interpretation, and (A3) for a constitutive one. However, not both readings are available: the ‘constitutive’ reading (A3) is not compatible with the Hypothesis, as it does not allow for violations that are still assertions. Scholars who claim that Williamson’s C-rule is constitutive in the orthodox sense are mistaken, because they incorrectly take (A3) to be equivalent to (A1). The regulative reading (A1), by contrast, is the original formulation introduced by the Williamson’s hypothesis, and hence an uncontroversial interpretation of the C-rule.

Importantly, saying that (A3) is not the C-rule falls short of claiming that (A3) is false. Even if (A3) is not the C-rule of assertion, it is a plausible definition of assertion *in terms of* its norm (A1). This definitional reading is compatible with the hypothesis, and it illustrates one of Williamson’s claims: that assertions can be defined as all and only the sayings that are subject to the C-rule.

It is now clear that the C-rule of assertion, as construed in the Hypothesis, is better interpreted along the lines of an orthodoxly regulative rule taking the form of (A1),an imperative-like norm that can be violated. In the next section, I will review interpretations that follow this insight, and construe the norm of assertion as an orthodoxly regulative rule possessing some salient ‘constitutive’ features.

## 4. The C-Rule as a Regulative Rule

### 4.1 The C-rule as orthodoxly regulative

In claiming that constitutive rules possess orthodoxly regulative properties (R2, R3), Williamson seems to treat constitutive rules as (orthodoxly) regulative. Hindriks (2007:396) draws the same conclusion from similar considerations (cf. also Maitra 2011:28f, McCammon 2014:137-9):

*In spite of the fact that Williamson* invokes the analogy with games, the knowledge rule cannot be a constitutive rule [...]. Constitutive rules specify (non-normative or descriptive) requirements an entity such as an action has to have in order to constitute another entity. [...] A related problem regarding the knowledge rule as a [...] constitutive rule is that [the rule] is a directive rather than (merely) a specification: it forbids assertions that do not express knowledge. Thus, instead of a constitutive rule, the knowledge rule is a regulative rule.

Nevertheless, there are good reasons to resist the conclusion that the term ‘constitutive rule’ is used to refer to a plainly regulative rule. First, it seems unreasonable to redefine the traditional notion of constitutive rule so that it is equivalent to its *complementary notion*, without even acknowledging this intention, and it would be uncharitable to attribute to Williamson this aim. Second, and relatedly, this interpretation diverges from how the majority of philosophers interpret the claim. Many authors either do not acknowledge that the norm is in fact orthodoxly regulative, or explicitly claim (e.g. Pagin 2015: §6.2, García-Carpintero 2014:143) that it is not. Rather than claiming that most authors are mistaken in their interpretation of the notion, it would be better to find an interpretation of the norm in the light of which it is possible to make sense of the existing debate.

### 4.2 Pollock’s paradigm: ‘prescriptive’ constitutive rules

A better prospect is to interpret the norm of assertion as a third kind of rule: a special kind of regulative rule that possesses some salient features of constitutive rules. After all, Williamson defines constitutive rules as possessing a feature (C1) of constitutive rules and features (R2, R3) of regulative rules.

More specifically, Williamson takes constitutive rules to possess (C1) in a slightly different sense, that I indicated with (C1\*). The main difference is that Williamson gives the notion a modal flavour: “*necessarily*, the rule governs every performance of the act”*.* Arguably, this modal relation is what makes the C-rule more significant than a mere regulative rule – one “intimately connected” to assertion (2000: 238). Hence, even if the C-rule possesses typical features of regulative norms, it is ‘constitutive’ of assertion in the sense that it *necessarily* regulates it: were assertion regulated by another norm, it would be a different speech act.

Before Williamson, Pollock(1982) had explicitly argued in favour of extending the label of ‘constitutive’ to all the orthodoxly regulative rules that *necessarily* regulate a practice –while acknowledging, unlike Williamson, his departure from orthodoxy. Pollock notes that these rules are such that “to eliminate [them] would be to profoundly alter the nature of games”, so that they belong to “the rules that *constitute* the nature of [those games]”, and hence “must be regarded as constitutive” (1982:212-213, italics mine). He consequently distinguishes two kinds of constitutive rules: orthodoxly constitutive rules (in his terminology, *definitive rules*), and regulative norms that necessarily regulate a practice (*prescriptive rules*). These two kinds of constitutive rules are opposed to merely regulative norms, which are not essential to the practices they regulate[[83]](#footnote-83). To sum up Pollock’s view:

* Constitutive rules:
* Definitive Rules: C1\*, C2, C3
* Prescriptive Rules: C1\*, R2, R3
* Regulative rules: R1, R2, R3

### 4.3 All norms are constitutive in Pollock’s sense

The notion of *prescriptive rule* seems to capture the sense in which the Hypothesis takes the C-rule to be constitutive of assertion: the C-rule is essential to assertion (C1\*), takes the form of an imperative (R2) and can be violated while engaging in the practice (R3)[[84]](#footnote-84). However, I will show that, depending on the understanding of ‘essential’ that one adopts, the notion of *prescriptive norms* either (i) fails to include the C-rule, or (ii) coincides with the orthodox notion of regulative rules. I will establish that Williamson’s Hypothesis is committed to interpretation (ii), and argue that this commitment comes with some problematic implications.

Williamson (2000:238-9) identifies two ways of conceiving a norm as *essential to* a practice. In a ‘broad’, ordinary sense, we say that a norm is essential to a practice if we cannot *conceive* the practice as not being ruled by that norm. This conception is quite loose, as it takes essentiality to depend on the judgments of a community. For instance, since in our community we say that “games such as tennis gradually change their rules over time without losing their identity” (2000:239), the rules that change in these games are not essential to them.

By contrast, in a ‘narrow’, technical sense, being essential is a very strict modal relation: the rules of a given practice define the identity conditions of that practice, so that they *necessarily* regulate it. Anychange in the rules of the practice generates a new practice identified necessarily by different rules, and in which different moves are allowed. In this technical sense, we say that a practice that evolves in time is a different practice at each stage of its evolution.

The C-rule cannot be taken to be essential to assertion in the *broad* sense, as this conception would not count it as constitutive of assertion (cf. Cappelen 2011:30). We obviously can conceive assertion as being governed by norms other than the C-rule: the very disagreement that animates the debate turns on the conceivability of different specifications of the C-rule: C could be that *p* is *true*, that *p* is *known* by the speaker, that the speaker *rationally believes p,* etc. The *broad* conception of essentiality is thus incompatible with the Hypothesis[[85]](#footnote-85).

Consistently, Williamson claims that the C-rule should be taken to be constitutive in the narrow sense (2000:239). The problem with this interpretation is that, paired with a ‘prescriptive’ interpretation of constitutive rules, it problematically erases the distinction between constitutive and regulative norms, counting all norms as constitutive – so that the claim that the norm is constitutive boils down to the trivial, uninteresting claim that the norm of assertion is a norm. Let us address this issue in more detail.

In the *narrow* sense, it seems that not only prescriptive rules, but also all the orthodoxly regulative rules are essential to (and hence constitutive of) the practices they regulate: every change in a regulative rule slightly alters the moves allowed in the practice, thereby generating a slightly different practice. To see this, consider an intuitively marginal rule of football:

(FKR) If a player kicks a free kick, she must not touch the ball again before it is touched by another player

It is easy to imagine variations of football that deny (FKR) and allow, e.g*.*, for a second or a third touch after the kick. In the ordinary *broad* sense, we judge such variations as unessential: a game without (FKR) can still be appropriately called football, and hence the rule is not essential to the game. However, in the technical, *narrow* sense that matters to the Hypothesis, (FKR) is essential to the game. Necessarily, (FKR) governs football: if football was regulated by a slightly different rule, football would be a slightly different game (call it football\*), identified by a slightly different set of rules.

If even a marginal regulative rule like (FKR) is counted as constitutive to football, it is difficult to see how this account could allow for rules that are *not* constitutive. It seems that every rule, independently of how marginal or central to a given practice, would count as constitutive on this conception. But if all rules are constitutive, claiming that a given rule is constitutive of assertion is redundant – it simply amounts to claiming that that rule is a rule of assertion.

Pollock (1982:218, cf. also Montminy 2014) attempts to address this worry by offering a putative example of a rule that would still count as regulative in the proposed framework: in the game of chess, “do not make unnecessary noise when your opponent is thinking about his next move”. Unlike (FKR), this rule is not essential to chess in the narrow sense; nonetheless, it regulates it. It is an example of a regulative, but not constitutive, rule of chess. Pollock’s account seems to classify general rules (like rules of etiquette) as regulative, and practice-specific rules as constitutive, thus maintaining a meaningful distinction between constitutive and regulative rules.

This explanation, however, overlooks a fundamental fact: while (FKR) is clearly a rule *of football*, the mentioned rule is not in the same sense a rule *of chess* (as a matter of fact, the first is in the rulebook of the game, the second is not). Rather than a rule of chess, the rule against disturbing the opponent is a general rule or maxim of a higher-order activity (for instance, a rule of *fair play*: “Avoid disrupting your opponent’s play, unless this is explicitly prescribed by the game”) that *also* constrains[[86]](#footnote-86), indirectly, which actions are appropriate in playing chess. To put it in other terms, the normative constraint against disturbing the opponent is not generated by the rules of chess, but rather by a higher order norm governing fair play. In a similar way, the norm “Do not speak with your mouth full” constrains the practice of assertion, but is not a rule *of* assertion. Rather, it is a general rule about how to eat and talk politely that also applies to the speech act of assertion.

If this line of reasoning is correct, the examples fail to show that the proposed account allows for some rules that are not constitutive. The aforementioned rules are constitutive too, because they are essential (in the narrow sense) to the higher-order practices that they regulate, namely ‘fair play’ and ‘eating politely’: if these latter practices were regulated by different rules, they would be different practices. However, for the claim that the norm of assertion is *constitutive* to be meaningful, one needs to show that there are some rules that are *not* constitutive. Otherwise, the claim that the norm is constitutive simply boils down to the trivial claim that the norm of assertion is a norm.

Perhaps there is way to reintroduce a constitutive-regulative distinction of some sort: even if all rules are ultimately constitutive, for any given rule it is possible to differentiate between the constraints it imposes *directly* (to the higher-order practice to which they are essential, *e.g.* eating politely) and the ones it imposes *indirectly* (to the lower-order practices to which they also apply, *e.g.* assertion), and dub the latter constraints regulative. In this sense, the *constitutive* rule of *eating politely* “Do not speak with your mouth full” is also indirectly a *regulative* rule of *asserting*.

On this reading, the distinction survives, but it is reduced to a mere matter of scope: all rules are ultimately constitutive, but when they impose constraint on a lower-order practice to which they are not essential, this constraint is of a *regulative* rather than *constitutive* kind. But is this the distinction that scholars writing on assertion have in mind, and is it significant enough to make the claim that the norm is constitutive significant?

### 4.4 The distinction is not significant

A first worry about the revised regulative-constitutive distinction is that it no longer identifies different *kinds of rules*, but rather different routes that a norm takes to constrain a given activity: a *direct* route (in Φ-ing, you are subject to the rule *qua* participant in practice Φ) or an *indirect* route (since in Φ-ing you are also X-ing, in Φ-ing you are subject to the rule, but in virtue ofparticipating in practice X). Thus revised, the constitutive-regulative distinction does not identify different species of rule, but instead different ways in which an activity can be subject to a rule of the same kind. Read in this light, the hypothesis that the norm is constitutive no longer specifies the kind of norm to which assertion is subject; it merely specifies in which way assertors are subject to a prescriptive norm: *qua assertors,* rather than *qua* participants in a higher-order activity.

A second, related worry is that this does not seem to be the role that Williamson (or any author writing on the norm of assertion) has in mind for his claim that the norm is constitutive. Williamson acknowledges that norms of higher-order practices can constrain assertions indirectly: “norms such as relevance, good phrasing, and politeness are just applications of more general cognitive or social norms to the specific act of assertion” (2000:238, cf. also 2000:256), but these remarks are independent from the claim that the norm is constitutive. Similarly, authors writing on the norm of assertion generally treat the claim that the norm is constitutive and the claim that the norm governs assertion directly as distinct.

In the light of these worries, someone defending a prescriptive interpretation could simply bite the bullet. Granted, one might say, specifying that the C-rule is *constitutive* tells us nothing about the nature of the norm (since all norms possess properties (C1-R2-R3)), but it still involves a meaningful specification, namely that the C-rule regulates assertion *directly*. And even though it seems that Williamson and other authors had a more significant role in mind for the claim that the norm is constitutive, this conception is still coherent and able to give *some* meaning to the Hypothesis.

At a closer scrutiny, defenders of the Hypothesis have *no other option* than biting this bullet, given the lack of viable alternatives. This is because there are only two alternatives on the table, and both unviable: to reject the revised distinction and accept the consequence that all norms are constitutive, in which case the claim that the C-rule is constitutive is reduced to the trivial claim that the C-rule is a norm; or to endorse the orthodoxly constitutive reading envisaged in section 4, which is incompatible with the Hypothesis.

The conclusion is that there is only one coherent interpretation available for the claim that the C-rule is constitutive of assertion. On this interpretation, the C-rule is a prescriptive norm in Pollock’s sense, and it differs from non-constitutive rules in that it regulates assertion directly, as opposed to indirectly. The upshot of this conclusion is that scholars writing on the norm of assertion have greatly overestimated the importance of claiming that the rule is constitutive. These authors have incorrectly taken the possession of properties (C1-R2-R3) to set apart norms that are different in kind from other norms, so that ascribing constitutivity to the C-rule was taken to ascribe a special status to the norm (e.g. Hindriks 2007, Bach 2007, Ball 2014). Having demonstrated that all norms possess properties (C1-R2-R3), this assumption has proven incorrect, forcing a retreat to the weaker claim that being constitutive of assertion (as opposed to non-constitutive) just means regulating assertion directly. As it will become even more apparent in the next section, this weaker notion of constitutivity is not significant enough to make the claim that the norm is constitutive interesting for the debate.

## 5. Conclusions

This chapter analysed a fundamental assumption underlying the debate on the norm of assertion: that assertions are subject to a *constitutive* norm, taking the form of the C-rule. It pursued two strategies to spell out the claim that the norm is constitutive. The first is to reinterpret the claim as orthodoxly constitutive, as suggested by some influential authors. This interpretation turned out not to be viable, as it is in plain contradiction with the hypothesis: no matter how we phrase it, a constitutive reading does not allow (by definition) for violations that are not assertions, which is against the hypothesis.

The second is to follow authors that interpret the norm as a regulative one with some constitutive features, namely as a ‘prescriptive norm’ in Pollock’s sense. This option revealed more promising, as it is consistent with the hypothesis: it allows for violations of the norm that are still assertions. However, authors have failed to realise that on this conception all norms that establish an obligation are constitutive, so that the claim that the norm is *constitutive* is reduced to the claim, usually taken to be different, that the norm regulates assertion directly.

While consistent with the hypothesis, this interpretation still has the disadvantage of being misleading. This is because the claim that the norm is constitutive is generally taken to imply that the norm a special status that sets it aside from other norms (Hindriks 2007, Bach 2007, Ball 2014) – a special status that the C-rule simply lacks (it does not differ from other norms setting an obligation, since these rules are also constitutive in the envisioned sense). The weaker claim that assertors are *directly* subject to a *prescriptive* *norm* avoids this misleading implication, and should therefore be preferred to the misleading claim that assertion is governed by a *constitutive* norm.

If assertion is not governed by a constitutive norm (or, to put it less bluntly, if a constitutive rule is just a rule that regulates assertion directly), what are the consequences for the ongoing philosophical discussion? The core of Williamson’s hypothesis still stands: it is still plausible that (i) assertion is subject to a norm taking the form: “One must: assert that *p* only if *p* has C” (i.e. that (A1) is true); and that (ii) this norm necessarily regulates assertion, in the Pollockian sense.

Other key claims and assumptions about the norm of assertion, however, lose their strength once we retreat to the claim that the C-rule regulates assertion directly. The first is the *uniqueness assumption*: the idea that assertion, *qua assertion*, is governed *only* by the C-rule. This assumption already rested on shaky grounds: it has been pointed out that “that there is such a unique rule is little more than an item of faith for Williamson, with no justification offered other than that a simple account consisting of such a single rule would be ‘theoretically satisfying’” (DeRose 2002:fn15), and several authors explicitly argued against this view (Brown 2008, Carter 2015 Carter & Gordon 2011, Gerken 2014, McKenna 2015).

The uniqueness assumption’s plausibility hinges significantly on conceiving the norm as constitutive in the sense of possessing a special set of properties, rather than in the sense of regulating assertion directly. This is because constitutive norms are regarded as unique in classic speech act theory: it is a standard view that every speech act is subject to a single constitutive norm and many regulative ones (Searle 1969). If the C-rule is constitutive in the orthodox sense, its uniqueness follows. But since the C-rule is not constitutive in the orthodox sense, there is no reason to assume that the C-rule is unique. And if assertion is not governed by a unique rule, it follows that (A3) is false:

(A3) One asserts that *p* iff in saying *p*, one is subject to the obligation that *p* must have C

Rather, (A3) should be replaced by:

(A4) One asserts that *p only if* in saying *p*, one is subject to the obligation that *p* must have C

Moving from (A3) to (A4) is not an insignificant step. It means abandoning the appealing idea that assertion can be defined simply in terms of its norm,that is, abandoning the claim that the norm is *individuating*. To be sure, the claim made here is not that (A3) must be false; rather, it is that once it is recognized that the norm is not constitutive, we have no warrant for preferring (A3) over (A4).

A further, crucial difficulty concerns the significance and originality of the hypothesis thus conceived. Deprived of the claims that the C-rule is constitutive, unique and individuating, the norm of assertion hypothesis is rendered uninteresting and of little significance. Even more importantly, this mitigated version is hardly a genuine hypothesis. It cannot be, because this view has already been orthodoxy in speech act theory for 50 years. It is a standard view that different speech acts are subject to different epistemic norms, and that these norms are analogous to the C-rule in form, *i.e.* that they are regulative rules in the orthodox sense (Searle 1969, Searle & Vanderveken 1985, Alston 2000).

In conclusion, all the innovative and interesting components of the hypothesis turn out to be the problematic ones that need to be abandoned: the assumption that the norm is unique and therefore individuating, and the introduction of a new conception of constitutive rules, that has led astray several scholars.

These observations have crucial consequences for the debate. In establishing that the norm of assertion is not constitutive, this chapter has not merely made a terminological point. It clarified that within this context ‘constitutive’ can only mean ‘prescriptive’, dispelling a great deal of confusion about the nature of the norm. Furthermore, it showed that it follows from this clarification that several assumptions in this debate are unwarranted – the *uniqueness* of the norm and its *individuating* role. The resulting, mitigated picture renders Williamson’s hypothesis of little significance. All things considered, we are better off abandoning the claim that assertion is governed by a constitutive norm.

# VI. Truth and assertion: rules vs aims

La parole a été donné à l'homme pou déguiser sa pensée

[Speech was given to man to conceal his thoughts]

Charles Maurice de Talleyrand

Now clearly, language, in its proper function, was developed not as a means whereby men could deceive one another, but as a medium through which a man could communicate his thought to others.

Augustine, Enchiridion de Fide, Spe et Charitatem, VII, 22

## 1. The disagreement about the norm of assertion

### 1.1 The norm of assertion

In the previous chapter, I have argued that the influential hypothesis that assertion is subject to a constitutive norm is subject to a number of difficulties. As a reminder, according to this hypothesis (Williamson 2000:241):

Williamson’s Hypothesis

There is just one rule. Where C is a property of propositions, the rule says:

(The C-rule) • One must: assert p only if p has C.

I have already rejected part of this hypothesis: I have shown that the claim that the C-rule is *constitutive* of assertion is misleading (since all rules are constitutive in the sense employed by Williamson), and questioned the claim that the C-rule is *unique* to assertions (*i.e.* that only assertions are uniquely subject to this norm), and that it *individuates* them (*i.e.* that assertions can be defined as the only speech acts subject to the norm). While these criticisms raise some important doubts over the solidity of the foundations of the debate, they do not dismiss Williamson’s Hypothesis completely. On the contrary, the idea that assertion is subject to *at least* one norm that regulates assertion and takes the form of the C-rule retains its plausibility. If we accept the hypothesis in this weakened version, a philosophically interesting question remains open, namely that of which property assertions should possess for them to be proper. In this chapter, I will review some attempts to answer this question, and develop my own proposal.

Several proposals have emerged in the literature, with different authors coming up with different ideas as to which norm regulates assertion – *i.e.,* as to which property C is. Broadly, they can be grouped into four categories[[87]](#footnote-87):

* (KR) Knowledge-Rule: “Assert *p* only if you know that p”

Williamson (1996/2000), DeRose (2002), Reynolds (2002), Hawthorne (2004), Schaffer (2008), Benton (2014), Turri (2010, 2011)

* (TR) Truth-Rule: “Assert *p* only if *p* is true”

Weiner (2005), Whiting (2012), Alston (2000), Wright (1992:§1)

* (JR) Justification-Rule: “Assert *p* only if you rationally believe that p” (or “assert *p* only if you are epistemically justified to believe that p”)

Douven (2006), Lackey (2007), Kvanvig (2009, 2011), Hill & Schechter (2007), Smithies (2012), Turri (2013b)

* (BR) Belief-Rule: “Assert *p* only if you believe that p”

Bach & Harnish (1979), Bach (2007, 2008), Hindriks (2007)

Of these proposals, the most demanding is (KR). Knowledge is understood to entail belief, justification and truth, so that in order to follow KR, you also have to follow all the other rules (JR, BR, TR). Some authors understand JR as requiring you to follow BR[[88]](#footnote-88), whereas no author takes BR or TR to entail any other rule.

Arguably, the starkest disagreement in this debate concerns whether a proper assertion needs to be true. Consequently, I will divide these four groups into two supergroups: those requiring truth for proper assertion (TR, KR) and those not requiring it (BR, JR). I will call the former accounts Factive, and the latter Non-Factive (since they do, or do not, require us to only state facts).

Factive Accounts: KR, TR

Non-Factive Accounts: BR, JR[[89]](#footnote-89)

### 1.2 Intuitions and norms

How can we determine which party is right about the norm of assertion? Games like chess, or institutionalised practices like witnessing in a courtroom (*e.g.* an Italian courtroom), are subject to explicit and well-defined rules. To determine which rules regulate these practices, one only needs to consult, respectively, the rulebook of chess and the (*e.g.* Italian) Penal Code. Linguistic rules, like the rule of assertion, are different: even if ordinary speakers enforce them, they are not collected in any ‘rulebook’, so that there is no text or authority to which we can refer to quickly settle the issue between factivists and non-factivists. Linguists would say that in these cases ordinary speakers have *procedural,* as opposed to *explicit* (or *declarative*)*,* knowledge of the relevant rules: we are able to follow them, but we do not normally have an explicit understanding of their nature (Chomsky 1965:4-8, Searle 1969:41-2 Mikhail 2011:19-21). The philosophical disagreement about the norm of assertion arises exactly from this lack of explicitness: since we only have procedural knowledge of the norm, enquiry is required to make this knowledge explicit.

Linguists tend to agree that even when we only have procedural knowledge of a linguistic norm, we will typically have intuitions about whether a relevant behaviour is in line with the norm or not. Applied to our inquiry, this means that we should have intuitions about whether any given assertion is appropriate or not *qua assertion*. For instance, suppose that I assert:

1. I want a beer right now

Almost all ordinary speakers will share the following intuitions about my utterance: that, *ceteris paribus,* assertion (1) is appropriate if sincere and true, and inappropriate if insincere and false. Intuitions of this kind are the linguistic data points against which theories about the norm of assertion are tested: a good theory should make predictions that are in line with the most widely shared intuitions, and avoid predictions that are at odds with them.

This also means that any account of the norm of assertion must live up to empirical evidence: its predictions cannot diverge too much from our linguistic practices (of praising and criticising assertions), because they must describe and make explicit these very practices, *i.e.* those in which our procedural knowledge of the norm of assertion is translated into action.

The assertion (1) does not present a challenge for any of the accounts that I introduced: every account is able to track the intuition that (1) is appropriate if sincere and true, and inappropriate if insincere and false. Since the focus of this section is the factive vs non-factive divide, my attention will be directed at cases that pull these apart ­– cases that are consistent with the intuitions of one party, but inconsistent with those of the other. In particular, I will focus on the phenomenon of Falsity-Criticism(that favours factive accounts), and on cases of Permissible Falsity(that favours non-factive ones).

### 1.3 Falsity-criticism and inadvertent violations

The primary aim of this chapter is to address the disagreement between factive and non-factive accounts of assertion. This is no easy task, because each account purports to explain a different set of equally plausible, but apparently incompatible, linguistic data. The main linguistic datum supporting factive accounts is what I call Falsity-Criticism. Falsity-Criticism can be spelled out both as an intuition and as a behaviour. It is a shared *intuition* that (*ceteris paribus*[[90]](#footnote-90)) a false assertion is bad and incorrect (and a true assertion is good and correct). And it is common *linguistic behaviour* that we praise assertions that match reality, and we criticise false assertions that diverge from reality.

Falsity-Criticism (and Truth-Praise)

* Linguistic Intuition*: false assertions are bad and incorrect in virtue of their being false (and true assertions are good and correct in virtue of their being true*[[91]](#footnote-91)*)*
* Linguistic Behaviour*: we criticise false assertions, and praise true ones*

Factive accounts can explain Falsity-Criticism (a label for both the intuition and the linguistic behaviour) in terms of conformity or lack of conformity to the norm, whereas non-factive accounts cannot (as noted in Williamson 2000: 262). To be sure, since false assertions are often unwarranted (or believed to be false), non-factive accounts will often be able to acknowledge that *in this subset of cases* a false assertion is wrong *in virtue of its being unwarranted,* (ordisbelieved)*.* However, this observation will only help in a subset of cases, so that Falsity-Criticism remains a problem for non-factive accounts. Furthermore, it seems that falsity constitutes a distinctive kind of wrongness. Any account failing to acknowledge that false assertions are incorrect *in virtue of their being false* is failing to explain a fundamental linguistic datum about assertion.

Proponents of non-factive accounts, however, deny the accuracy of this linguistic datum. They point out that *inadvertently* saying something false (*i.e*. making an *unlucky assertion*) does not generally amount to bad linguistic behaviour, and *inadvertently* saying something true (*i.e*. making a *lucky assertion*) does not generally amount to good linguistic behaviour (we will see this in more detail in the next section). In fact, we do not criticise *unlucky* assertions, and we do not praise *lucky* ones. In other words, proponents of non-factive account purport to explain a different set of linguistic data:

Permissible Falsity (and Impermissible Truth):

* *Linguistic Intuition: inadvertently false* (‘unlucky’) *assertions are permissible; inadvertently true* (‘lucky’) *assertions are impermissible*
* *Linguistic Behaviour: we do not criticise unlucky assertions, and do not praise lucky ones*

In this chapter, I will address the following puzzle. Both Falsity-Criticism and Permissible Falsity appear to be accurate descriptions of our intuitions and of our linguistic behaviour. Only factive accounts are able to explain Falsity-Criticism, and only non-factive accounts can explain Permissible Falsity. But these two sets of linguistic data (and the respective accounts of assertions) are incompatible: according to the former, false assertions are always impermissible; according to the latter, they are permissible under some circumstances.

This is my plan to solve the puzzle. Section 2 introduces and discusses (un)lucky assertions, showing that it is not easy to dismiss this linguistic datum. Section 3 argues that, against what is generally taken for granted, Falsity-Criticism is compatible with a non-factive view of assertion. Yet, proponents of non-factive accounts have never provided an explanation of Falsity-Criticism in non-factive terms. Section 4 develops this explanation: I argue that truth can be understood as the aim, rather than the rule, of assertion. This yields a non-factive explanation of Falsity-Criticism that still allows for Permissible Falsity. Section 5 deals with some possible objections to my view.

## 2. Lucky and unlucky assertions

### 2.1 Lucky: inadvertent observation of TR

Within the factive family, the truth-rule (TR) offers the simplest explanation for Falsity-Criticism. We should only assert what is true: conformity to TR elicits praise, and its violation criticism. Despite its appeal and simplicity, however, TR is too permissive: ‘lucky assertions’ offer a first linguistic data point against this account. Lucky assertions are true assertions that the speaker does not believe to be true. It seems that these assertions are not proper, despite their compliance with TR. To see this, consider the following example:

Lucky Assertion[[92]](#footnote-92)

[Speaker follows TR, violates BR]

Giacomo and Frida are married, but Giacomo is looking for an affair. At a party, he starts flirting with a woman called Giuseppina. Giuseppina knows that he is married. Deviously, Giacomo decides to tell her:

1. My wife has been cheating on me in the last months

Giacomo is lying: he is trying to make Giuseppina feel ok about flirting with a married man. As it turns out, unbeknownst to Giacomo, his wife Frida was really cheating on him.

Intuitively, (1) is not appropriate *qua assertion*: Giacomo lied to Giuseppina. The fact that, unbeknownst to Giacomo, Frida was cheating on him does not make his assertion appropriate. TR is not able to make this prediction: it predicts instead that (1) is a perfectly appropriate assertion. This constitutes a first data point against TR.

### 2.2 Unlucky: inadvertent violation of TR/KR

Lucky assertions pose a challenge to TR, but not to the knowledge-rule (KR): KR correctly predicts that (1) is inappropriate, since Giacomo does not know that what he said is true. For a counterexample to factive accounts *in general* (*i.e.*, both TR and KR) we need to consider “unlucky assertions”: false assertions that the speaker reasonably believes to be true.

Let us consider an example based on Ginet’s famous fake-barn scenario:

Unlucky Assertion

[Speaker follows JR and BR, violates TR and KR]

Lorenzo and Jacopo are driving in the Tuscan countryside. They are driving through a hilly area, with barns on top of each hill. Lorenzo declares: “I’d like to take a photo of a yellow barn”, to which Jacopo replies:

1. There is a yellow barn on our left

Unbeknownst to both, they are driving through fake barn county, an area where almost all barns have been substituted by barn facades (facades that, seen from the distance, are undistinguishable from real ones). On their left lies one of the many fake barns, not a real barn.[[93]](#footnote-93)

Jacopo’s assertion (2) is false, so that it constitutes a violation of both KR and TR. Most people share the intuition that, given the available evidence, it is appropriate for Jacopo to assert there is a barn on his left. This intuition is generally justified as follows: Jacopo has perceptual evidence to support his belief that (2) is true, and no good reason to doubt the reliability of his perceptions; since any responsible speaker in his position would have asserted the same proposition, it seems counterintuitive to judge his assertion inappropriate, or his linguistic behaviour criticisable. If this line of reasoning is correct, the non-factivist argues, unlucky assertions constitute a counterexample to both TR and KR, and these accounts should be abandoned in favour of a non-factive account of assertion.

To object to the relevance of this counterexample, one may reply that linguistic rules are meant to deal with the vast majority of cases, and that they can allow for some exceptions of little significance. If unlucky assertions are only uttered in very rare and highly unusual situations, they pose a negligible challenge to factive accounts, and they can be discounted as a scarcely relevant exception. It seems easy to argue that unlucky assertions are exceptional. In the majority of cases, when I reasonably believe that *p*, *p* is also true. Scenarios like ‘fake barn county’, by contrast, are highly unusual: it is extremely unlikely that our perceptions get misguided in this way. And this is the point of the example somehow – was the fake barn scenario not an exception, Jacopo would not be epistemically justified in trusting his senses. Hence, since unlucky assertions are by definition exceptional cases, they can be ignored when it comes to determine which norm governs assertion – *quod erat demonstrandum*.

This conclusion, however, is too hasty: unlucky assertions are not as exceptional as ‘fake barn’ scenarios. An unlucky assertion is an assertion that *p* uttered while holding a justified false belief (JFB) that *p*. As any philosopher well knows, cases of JFB are not limited to perceptual mistakes like the fake barn scenario, and actually involve quite mundane cases of epistemic misfortune. We are all familiar with *getting things wrong*: sometimes it is our fault, sometimes it is not. The latter situation is the one involved in unlucky assertions: we hold a JFB when we get things wrong and it is *not* our fault – when we get them wrong *despite* the fact that we base our belief on good reasons. Consider some everyday examples of beliefs that can be false despite being supported by good reasons: I believe that the water that is now coming out of my tap is not poisoned because it has never been before; that the economy is recovering because it was reported consistently on several media; that my favourite necklace is in my safe because I remember putting it there. In all these examples, I could be wrong *despite* having good reasons supporting my belief – inductive, testimonial and mnemonic reasons respectively. Were I to assert any of these propositions, my assertion would be an unlucky one; but unlike the fake barn scenario, these cases are mundane and familiar, showing that unlucky assertions (asserting a JFB proposition) can hardly be dismissed as exceptional, deviant cases.

No matter how careful we are, we are fallible epistemic enquirers, and we sometimes end up forming JFB, and consequently asserting falsities – falsities backed up by reasons; but, unbeknownst to us, the wrong reasons. Unlucky assertions cannot be dismissed as irrelevant exceptions, and represent a substantive problem for factive accounts. Nonetheless, a factivist philosopher could always deny the intuitions that unlucky assertions are impermissible: in the next section, I will show that such a claim leads to controversial conclusions[[94]](#footnote-94).

### 2.3 Checkmating the speaker: a too demanding norm

He who knows, does not speak.

He who speaks, does not know.

Laozi, Tao Te Ching (chapter 56)

When exposed to unlucky assertions, some factivists are willing to bite the bullet, and insist that unlucky assertions are not permissible, because they are false. The onus now is on them to provide a plausible story about the alternative course of action that they would recommend to unlucky speakers: what are they supposed to assert instead? Going back to the barn example, if Jacopo’s assertion is impermissible, what is he allowed to assert in response to Lorenzo’s request to show him a yellow barn? Defenders of factive accounts have failed so far to provide a convincing answer to these questions – as a matter of fact, they dodged it altogether, insofar as scenarios involving assertions *prompted by a question* are not discussed in the literature. To see if they can nonetheless provide a convincing response, I will consider what a factivist *could* say in this respect.

According to TR, in the proposed scenario Jacopo can only assert something true. A negation of (2), *i.e.* ¬(2) (there is no yellow barn on the left), would satisfy TR. But recommending the assertion ¬(2) is counterintuitive. Jacopo would be insincere about the position of the barn, and only inadvertently saying something true: in other words, he would be making a lucky assertion, and we have established that lucky assertions are impermissible. The prediction of TR is clearly off track.

What does KR recommend? There is nothing that Jacopo knows about the whereabouts of yellow barns in his vicinity, so that KR seems to only allow for one course of action: to stay silent (an option that, on closer inspection, is also permitted by TR). However, it would surely be unacceptable linguistic behaviour for Jacopo to stay silent in response to Lorenzo’s question. This is an undesirable prediction: any theory that leaves a speaker checkmated into silence in a situation of this kind cannot be a good theory of assertion.

Perhaps there is another recommendation that is compatible with both KR and TR: that Jacopo should make a hedged assertion, like “It appears to me that there is a yellow barn on our left” or “I believe that there is a yellow barn on our left”. This prudential mitigation seems an appealing solution, as it prevents Jacopo from saying something false, while not requiring him to say something he disbelieves or does not know. However, this is not a plausible recommendation either, as it could not generalise coherently to our linguistic practice. By definition, we do not know when we hold a justified false belief about a proposition, so that by definition we cannot know when we are about to utter unlucky assertions. If Jacopo were to mitigate his assertion with an evidential (e.g. ‘apparently’, ‘it seems to me that’, etc.), the appropriateness of the mitigation would be a matter of luck, rather than a matter of responsible linguistic behaviour: his assertion would be, once again, a lucky assertion. More generally, a linguistic rule recommending to mitigate a subset of assertions that by definition speaker cannot identify is boldly implausible.

Another way to abide to factive standard would be to *always* mitigate assertions, thereby eliminating the possibility of unlucky assertions. But also this requirement is implausible, as it would result in a language in which every assertion is introduced by evidentials – and in which evidentials would eventually serve no genuine mitigating function, as all assertions would be mitigated.

The impossibility of replying in any way when questioned about a proposition we falsely think we know illustrates the counterintuitive component of factive accounts, namely their excessive demandingness. These accounts recommend a linguistic conduct that is too distant from our own – worryingly close to that of the ‘wise’ from the *Tao Te Ching* from the quote above, who observes the knowledge rule so strictly that he ends up asserting very rarely (or, as the hyperbole in the passage recites, not asserting anything at all).

## 3. Deriving permissibility from correctness

Even though these arguments are convincing, the non-factive account of unlucky assertions may still seem incomplete. There is a sense in which, after all, unlucky assertions are *not* correct. Consider the fake barn example again. Even if Jacopo is justified in asserting “There is a barn on our left”, competent speakers will agree that his assertion is incorrect in some important sense: namely, in the sense that it is false. This observation is in line with the Falsity-Criticism intuition: a good account of the norm of assertion should acknowledge that false assertions are incorrect. We are faced once again with the puzzle introduced at the beginning of the chapter: unlucky assertions are intuitively permissible (as for Permissible Falsity), and still there is *a sense* in which they are incorrect and wrong (as for Falsity-Criticism intuition). Each account is supported by equally plausible intuitions, but these intuitions seem incompatible.

I want to argue that the way out of this puzzle lies in identifying the relevant meaning of the words ‘incorrect’ and ‘wrong’. In what follows, I will show that ‘incorrect’ and ‘wrong’ do not always indicate that an agent is in violation of a norm: an action can be incorrect and nonetheless in compliance with a relevant rule. If this applies to unlucky assertions, we cannot derive their impermissibility from their incorrectness: there is no contradiction in claiming that they are both *permissible* and *incorrect*, and thus no contradiction between Permissible Falsity and the Falsity-Criticism intuition.

Let us start by reconstructing the factivist argument deriving the impermissibility of false assertions from the Falsity-Criticism intuition – more specifically, from the ‘incorrectness’ of false assertions. The following argument is taken almost literally from Whiting (2012: 848):

Argument from Correctness:

1. It is a platitude that:

(C) One correctly asserts that *p* only if it is true that *p*

1. Statements about what it is (in)correct to do entail statements about what one may (not) do[[95]](#footnote-95)
2. Given 2, from (C) we can derive (TR):

(TR) One may: assert that *p* only if it is true that *p*

The Argument from Correctness attempts to establish TR (*i.e.* factive accounts) starting from (C), a version of the Falsity-Criticism Intuition. (C) can be read as equivalent to the claim that false assertions are incorrect (at least in the framework of bivalent logics), a claim that I accepted as part of the Falsity-Criticism Intuition. Premise (2), according to which judgments of incorrectness are based on normative judgments about what one is allowed to do, is more controversial, and can be easily challenged.

To be sure, (2) is often true: in many cases, judgments of correctness track judgments about permissibility. For instance, I can say that your behaviour at my brother’s marriage was *correct* for that kind of event, meaning that it was appropriate, that it complied with the rules of social behaviour that apply in those contexts. However, at least in some contexts, ‘correct’ can simply mean ‘true’. And if correct *can* mean true, then not every judgment of correctness tracks a judgment about permissibility, against premise (2).

Most dictionaries will support this observation. According to the OED, for instance, the primary meaning of ‘correct’ is:

(M1) Free from error; in accordance with fact or truth

The OED also lists some secondary meanings of ‘correct’ that are more aligned with the Argument from Correctness. While (M2) is still close to (M1), (M3) and (M4) are compatible with the interpretation of (C) proposed by the factivist – namely, that correctness indicates permissibility:

(M2) Not mistaken in one's opinion or judgement; right:

(M3) Meeting the requirements of or most appropriate for a particular situation or activity

(M4) Conforming to accepted social standards; proper[[96]](#footnote-96)

Which meaning of ‘correct’ is employed in the Argument from Correctness? Premise (1) states that it is a *platitude* that (C), *i.e.* that “one correctly asserts that *p* only if it is true that p”. (C) is clearly not a platitude if we interpret ‘correct’ as ‘permissible’: whether asserting some falsities can be permissible is after all one of the main questions that animates the debate on the norm of assertion. (C) is clearly a platitude, by contrast, if we interpret ‘correct’ as in (M1-2), that is, as indicating truth: this is unquestionably the sense employed in (1).

Opposite considerations apply to premise (2). Premise (2) is true only if ‘correct’ means ‘permissible’, as in (M3-4): judgments about permissibility are indeed judgments about what one may do, whereas judgments about the truth of a proposition (as in M1-2) are clearly not judgments about what one may do. Premise (2) relies then on the other sense of ‘correct’, the normative sense. The Argument from Correctness is then invalid, as it relies on an equivocation between two different meanings of ‘correct’: (M1-2) in premise (1) and (M3-4) in premise (2).

Perhaps, however, the Argument from Correctness fails to represent the factivist line of argumentation. To argue that unlucky assertions are incorrect, all one needs is a revised, weakened version, that avoids these objections and establishes the same conclusion. First, the factivists can maintain that in premise (1) ‘correct’ can and should be interpreted as ‘permissible’, as in (M2-M3), thus avoiding the accusation of equivocating between different meanings[[97]](#footnote-97) . Second, to allow for the first move, they can weaken premise (1) by dropping the claim that (C) is a *platitude*: they only need to accept that it is *intuitive* that (C) is true.

These revisions will not lead very far. We have already seen that premise (1) is controversial if ‘correct’ is interpreted as ‘permissible’. This is true even by mitigating (C) as merely intuitive rather than as a platitude: the intuition that assertions are *permissible* only when they are true is very weak, it is challenged by unlucky assertions, and cannot be granted as a straightforward premise. More importantly, that assertions are permissible only if true is exactly what the argument is trying to establish. If ‘correct’ is interpreted as ‘permissible’ in (1), the argument is reduced to a brazen *petitio principii*: since only true assertions are permissible, it is only permissible to assert true propositions.

In conclusion, the factivist cannot really produce an argument to derive impermissibility from incorrectness[[98]](#footnote-98). We cannot straightforwardly draw conclusions about the norm of assertion from the Falsity-Criticism Intuition. This means that the puzzle envisaged at the beginning of this chapter is solved: once we distinguish between incorrectness and impermissibility, it is uncontroversial that unlucky assertions can be both *permissible* (satisfying Permissible Falsity) and (in some sense) *incorrect* (satisfying Falsity-Criticism).

## 4. Truth as aim

There are a thousand ways of missing the bull's-eye,

only one of hitting it.

Montaigne, On Liars (Essays, 1.9)

So far, I have established that Impermissible Falsity represents a genuine challenge for factive accounts. Furthermore, it turned out that the Falsity-Criticism Intuition does not necessarily have implications for the norm of assertion, as we cannot derive impermissibility from the incorrectness of false assertion. While this suggests that non-factive accounts are better suited to explain the linguistic data points introduced in section 1, there is still some room for manoeuvre for the factivists to defend their cause. The factive account still has a trick up its sleeve: the Falsity-Criticism Linguistic Behaviour, introduced the first section:

Linguistic Behaviour: we criticise false assertions (and praise true ones)

Factive accounts are able to explain this phenomenon in terms of compliance or disobedience to the norm. By contrast, non-factive accounts lack any explanation for this phenomenon. So far, no non-factivist author has yet addressed this problem, providing a non-factivist story as to why assertions are praised for being true and criticised for being false. Absent such explanation, factive accounts maintain some appeal: they can explain Falsity-Criticism, while non-factive accounts cannot. In this section, I will attempt to fill this gap, and show that such explanation can be offered within a non-factivist framework: as anticipated, my proposal is that truth is the purported aim of assertion, and that criticisms and praises can be explained in terms of success and failure to meet the aim of assertion.

### 4.1 The difference between rules and aims

The view that assertion is governed by norms is typically conceived “by analogy with the rules of a game” (Williamson 2000:239, cf. also Kölbel 2010). An often-undervalued point is that games not only have rules, but also (purported) aims: aims that a player purports to have in playing the game (cf. Schwyzer 1969, Maitra 2011) And there is an intuitive difference between the rules of a game and the aims of a game.

Arguably, the essential ‘aim’ of competitive games is that of winning (e.g. Kemp 2007: 113): in this sense, checkmate is the aim of chess. The difference between aims and rules is quite intuitive in games: even if the aim of chess is to checkmate, it is not a rule of chess that you should *only* make moves that lead you to checkmate. Now, often competitive games also involve derivative, intermediate aims – things that you standardly aim to do in order to win the game. For instance, the aim of shooting a penalty in football is to score a goal, because only scoring a goal will lead you to win the game. These intermediate aims are also clearly not rules: while scoring is the purported aim of shooting a penalty, it is not a rule of football that you should shoot the penalty only if you score a goal.

These examples show that there is an intuitive difference between the aims and the rules of a game. To offer a broad, general characterisation of the difference between aims and rules, we could define them as follows:

**Difference between Rules and Aims**

For every condition X, action-type φ, and action-token φ1:

* A is the (only) aim of φ iff φ1 is successful if (and only if) A
* R is the (only) rule of φ iff φ1 is permissible (if and) only if R

Let us consider two examples from football to illustrate this. Saying that *scoring a goal* (A) is the aim of *shooting a penalty* (φ) means that any given *instantiation of penalty-shooting* (φ1) in football is successful if *a* *goal is scored* (*i.e*. condition A is met). By contrast, saying that *not touching the ball with the hands* (R) is a rule of *ball-touching* (φ) in football means that any given *instantiation of ball-touching* (φ1) in football is permissible only if *the ball is not touched with the hands* (*i.e.* condition R is met). The difference here is that the aim identifies a condition for *successful* action, whereas the rule identifies conditions for *permissible* action.

While aims and rules are different, they are also similar in some respects. One important similarity is that they can both ground evaluative judgments in the right circumstances. We can *criticise* people for failing to fulfil aims as well as for violating rules, and we can *praise* them both for succeeding to fulfil certain aims and for following rule (more on this in section 5.4). This suggest that there is potential for both aims and rules to ground evaluative judgements – and therefore potential for a putative truth-aim of assertion to account for the Falsity-Criticism Behaviour.

### 4.2. Truth as a rule, truth as an aim

Two explanations are now available to explain the Falsity-Criticism Behaviour: a factive explanation (truth is the *rule* of assertion), and an explanation that is compatible with non-factive accounts (truth is the *aim* of assertion). The latter explanation has a great advantage over the former: it enjoys the same explanatory power, while avoiding the problem of regarding unlucky assertions as impermissible. Nonetheless, it is yet to be proved that truth is the aim of assertion: we need positive reasons in support of this hypothesis.

A fairly strong case in favour of a truth-aim account can be made by going back to the distinction between rules and aims. Can we develop a test to tell apart rules from aims, a test useful in situations in which the rules and aims of an activity are not explicitly formulated, as in the case of the norm of assertion? For this purpose, we need to look at the most obvious differences between aims and rules. One is the following: aims characteristically allow for unintentional failure, whereas rules do not typically allow for unintentional failure. In other words, typically aims are such that you can try but fail to meet them, whereas rules are such that you cannot easily try but fail to follow them.

To see this, consider some examples of aims: in shooting a penalty, I can fail to score a goal despite my best effort; similarly, in playing chess I can fail to checkmate even if I play to the best of my capabilities. The same is not the case for rules: it is difficult for a *competent and careful* player to try but fail to move the bishop only diagonally, or to try but fail to castle only if the king and rooks have not yet moved.

These observations suggest that we could accept something like the following test for distinguishing a rule-based activity from an aim-based activity:

Test for the rule-aim distinction:

*For any given action A within a social practice P, for any given condition C applying to A, for any competent and careful agent S,* the truth of (T) is *prima facie* evidence that C is the aim of A in P, whereas the falsity of (T) is *prima facie* evidence that C is the rule of A in P:

(T) S can typically try but fail to meet C

The test only specifies a condition for *prima facie* evidence that C is the aim of A: I do not need to commit to a stronger version, as for my purposes the test can allow for some countervailing considerations. The adverb ‘typically’ occurring in (T) requires some specification: broadly, I take the possibility of failure to be typical if it is a live possibility also for competent and careful agents. For instance, an incompetent or careless player of chess may fail to castle according to the rule, whereas a competent and careful player may not do so.

Is the test valid? Let us consider some apparent counterexamples. One is the offside rule of football: it seems that this rule is often violated inadvertently by players that nonetheless are trying to follow it. However, such players cannot be described as being careful: they failed to check the line of defence, either carelessly (*e.g*. they got distracted by the chance to score) or purposefully (*e.g*. sacrificing mindfulness to focus on the action). Admittedly, carefulness is a matter of degree, so that the Test will allow for a few borderline cases; nonetheless, this is enough to treat correctly the vast majority of cases.

A more telling apparent counterexample is that of the rule against touching the ball with your hands. It is fairly easy to touch the ball with your hands inadvertently in football, so that the rule seems to constitute a counterexample to the Test. However, the actual football rule only prohibits “a *deliberate* act of a player making contact with the ball with the hand or arm” (IFAB rulebook 2016/2017:82), so that you cannot try but fail to follow it.

This rule is interesting because it regulates an action that can be easily performed inadvertently. Given the high risk of inadvertent hand-touch, a rule prohibiting each and every hand-touch would be perceived as unfair – as it would often cause players to be punished for actions they performed inadvertently. To avoid this, the rule is purposefully designed to punish only deliberate violation. I am willing to speculate that a similar point applies to most rules: whenever there is a relevant risk of inadvertent violations, rules tend to be designed to prohibit only intentional violations, because the opposite would be unfair.

These observations are in line with the Kantian principle that ‘ought implies can’ (Kant CPR). On a strong interpretation, this principle postulates that it is conceptually impossible for you to be morally required to do something that you are not able to do. On a much weaker interpretation, this principle can be interpreted as hinting that it is unfair to require someone to do something he is not able to do – we may call this the ‘ought ought to imply can’ principle, to indicate the mitigation of the Kantian original. While there is disagreement as to whether the strong version of the principle holds (Ryan 2003, Stern 2004, but cf. Kurthy, Lawford-Smith & Sousa 2017), there is very little ground to criticise its weak version: when it comes to rules (of a game, of a social activity), it is in an important sense unfair to establish rules that participants are not able to follow, that they are often unable to follow, or that they can easily violate inadvertently.

If something like ‘ought ought to imply can’ holds for rules, it can explain the rationale behind the proposed Test. Rules typically do not allow for unintentional failure because, according to ‘ought ought to imply can’, it would be unfair to do so; aims are not subject to the principle, and so they can allow for unintentional failure. The test is thus not merely motivated by a statistical observation about the frequency of rules and aims that allow for unintentional failure, but also by independentconsiderations over intuitive constraints of fairness that apply to rules governing social practices.

Overall, there seems to be a strong case in support of the Test, and the Test tips the balance heavily in favour of the truth-aim account. The truth-*rule* account untypically and unfairly allows for inadvertent violations (like unlucky assertions) and inadvertent compliances (like lucky assertions). The truth-*aim* account, by contrast, typically and fairly counts unlucky assertions as unsuccessful but permissible, and lucky assertions as successful but permissible. The Test tells us that truth is the aim, rather than the rule, of assertion.

### 4.3 The truth-aim account

I have identified strong reasons to believe that there is a sense in which truth is the ‘aim’ of assertion. But this claim is still very vague, as it is not obvious in which sense a speech act like assertion can aim at a truth-value. My proposed view is that this means that an assertion is successful only if it is true, and this is because assertions conventionally have a *word-to-world* direction of fit. I will explain both claims, and outline some connections with existing scholarship.

The first claim derives from the proposed account of aims, according to which “X is the (only) aim of φ = φ1 is successful if (and only if) X”. On this conception, saying that assertion aims at truth means that an assertion is successful only if it is true, more or less in the same way in which saying that the aim of a penalty is to score a goal means that the penalty is successful only if you score a goal. To be sure, the relevant standard for successfulness is the one set by the relevant practice that we are considering, not by the agent. For instance, in playing chess, I may aim to bore my opponent, thereby preferring a stall to a victory. In succeeding in stalling, I would be successful with respect to my personal aim, but not with respect with the purported aim of playing chess. Similarly, in asserting I might have the personal aim to say something false, and be successful with respect to that aim, not with respect to the purported aim of asserting.

We need to differentiate between the personal aims that one might have in engaging in a practice and the aims that are set by the practice in which one engages. To stress this distinction, I call the aims that are determined by the practice the *purported aims*[[99]](#footnote-99)of that practice, and any divergent aim (whether selfish or not) a *personal aim*. In this sense, checkmating is the *purported aim* of chess, scoring is the *purported aim* of shooting a penalty, and saying something true is the *purported aim of assertion,* whereas the divergent aims from the previous examples count as *personal aims*. Once this distinction is clear, it becomes apparent that considerations about personal aims are not a challenge for my view.

As for the second component of my explanation, that assertions have a *word-to-world* direction of fitis anything but a controversial claim – one that I already defended in chapter III (section 3.1). It is standard to read the notion of direction of fit as having a teleological connotation, *i.e*. as setting the condition of success for a speech act, rather than the conditions for its appropriate performance (Humberstone 1992). As a matter of fact, often authors treat ‘having truth as an aim’ (understood as having truth as a success condition) and ‘having a *word-to-world* direction of fit’ as synonymous: “’True’ and ‘false’ are the favourite terms of assessment for achieving *success* of illocutions which have the word-to-world direction of fit” (Searle 2007: 34).

As already mentioned in chapter III, speech act theorists agree that not *only* assertions have word-to-world direction of fit. This is an important reminder. Several speech acts aim at truth, namely all speech acts belonging to the class of ‘assertives’ (cf. Searle 1976, also called ‘representatives, or ‘constatives’, as in Recanati 1981): conjecturing, hypothesising, testifying, objecting are but a few examples. The success condition for all these speech acts is that their propositional content is true: a conjecture that *p* is successful if *p* is true, the same for hypotheses, objections, etc. The bottom line is that purportedly aiming at truth is a necessary, but not sufficient, condition for an utterance to be an assertion: the truth-aim account explains Falsity-Criticism, but does not provide a definition of assertion on its own.

Overall, my proposal to understand aiming at truth as a success condition determined by the direction of fit can claim a strong link with an orthodox view in speech act theory[[100]](#footnote-100), and promises to be able to solve the puzzle presented in the opening of this chapter. Unlike factive accounts, it is compatible with Impermissible Falsity; unlike existing non-factive accounts, it explains the Falsity-Criticism Behaviour and why we deem true assertions ‘correct’: not because they follow the rule of assertion (as the Argument From Correctness holds), but because they ‘fit’ reality in the right direction. Furthermore, it finds independent motivation from the Test for the Rule-Aim distinction.

### 4.4 Ought to try – from aims back to rules

The truth-aim account of assertion is by definition non-factive. Rules and aims, as I defined them, are mutually exclusive: if truth is the aim of assertion, it cannot be its rule. But the truth-aim account of assertion is compatible with *any* non-factive account: any non-factive norm (JR, BR, etc.) can be paired with TA. This elicits the question: which particular non-factive norm should be paired with TA? For the purpose of this thesis, I do not wish to commit to a definite response to this question: my aim was only to show that the puzzle I presented can be explained by pairing a non-factive account with the view that assertion aims at truth.

Nonetheless, I would like to conclude with a suggestion: if assertions aim at truth, it seems intuitive that asserting cooperatively requires one to *try to say the truth*. This idea follows from a more general consideration: that cooperating towards a given aim seems to require cooperating agents to try to achieve that aim. To see this, suppose I am helping you to find your way home, a joint cooperative action that has your finding the way home as a success condition. Since the purported aim of our actions is to get you to find the way home, it seems that, to cooperate in helping you to find your way home, I must try to make you find the way home. If cooperating towards an aim requires one to try conform to that aim, since assertion aims at truth, to assert cooperatively one should try to say the truth. In other words, we can infer the non-factive ‘truth-aim rule’ of assertion (TAR) from (TA) and (CP):

(TA) Truth is the aim of assertion

(CP) To cooperate in a teleological activity you must try to conform to the purported aims of that activity

(TAR) You should: assert *p* only if you are thereby trying to assert a true proposition

In inferring TAR from TA-CP, I am smuggling in a further assumption: that assertion is (or purports to be) a cooperative action. While this view is fairly standard[[101]](#footnote-101), it can certainly be criticised. Nothing, however, hangs on this assumption. Perhaps TAR cannot be derived from TA, or perhaps it can be derived with a different strategy. But it is undeniable that there is an intuitive link between TA and TAR. If the derivation does not hold, I am still willing to suggest that if one accepts TA, TAR is an obvious candidate as a norm of assertion.

Versions of TAR have been defended by influential philosophers before the debate on the norm of assertion came into fashion. Dummett endorses it almost literally: “Assertions are governed by the convention that we should try to utter only [true ones]” (Dummett 1981:356). Lewis’ definition of truthfulness as a rule that is constitutive of a language £ can be interpreted as a version of TAR: “try never to utter sentences of £ that are not true in £”. Similarly, Grice (1989) has defended (TAR) as a maxim governing saying something in cooperative communication.

Grice’s position comes with a further suggestion. In his view, the ‘Quality Maxim’ TAR can be unpacked into two submaxims. These two submaxims are strikingly similar to the two most influential non-factive accounts on the market: the belief-rule (BR) and the justification-rule (JR). In other words, if we follow Grice, we can further derive BR and JR from TAR:

Quality Maxim: “Try to make your contribution one that is true” (TAR)

Submaxim #1: “Do not say what you believe to be false” (BR)

Submaxim #2: “Do not say that for which you lack adequate evidence” (JR)

Admittedly, there are several important differences between the Gricean maxims and JR and BR[[102]](#footnote-102). For our purposes, however, we can leave aside of whether this interpretation really does justice to Grice’s own view. The relevant question is: can we derive BR and JR from TAR? The short answer is no. There are speech acts that have truth as an aim like assertion (speech acts with *word-to-world* direction of fit), that arguably are also subject to TAR, but that are subject to weaker norms than BR or JR. Hypothesising and guessing, for instance, have word-to-world direction of fit: they are successful if they are true. They are intuitively subject to TAR, but not to additional norms as strict as JR/BR. It seems that I can appropriately hypothesise that *p* even if I do not yet believe *p*, and that I can appropriately guess that *p* even if neither believe that *p* nor have good evidence for *p.* The norms of hypothesizing and guessing could be something along the following lines:

(ER) Hypothesize that *p* only if you have evidence that *p*

(KFR) Guess that *p* only if you do not know that *p* is false

Whether hypothesising and guessing are subject to these *exact* rules is not relevant: as long as they are subject to norms that are different from BR/JR, they represent a counterexample to the ‘Gricean’ conjecture.

Still, a grain of truth can be extracted from the conjecture. If all assertives are subject to TAR, they cannot be subject to just about any rule: they can only be subject to rules consistent with TAR. Consider again my examples. The rules JR and BR, that many authors believe to constrain assertion, are clearly compatible with TAR. ER and KFR are also compatible with TAR, but less demanding. Clearly there can also be stricter rules, as in the following putative rule for testifying that *p*:

(FPR) Testify that *p* only if you experienced that *p* in first person

The emerging picture is that there is a family of speech acts, assertives, whose members aims at truth. All these speech acts have a normative constraint in common: TAR. But they are also subject to additional epistemic constraints (some stricter, some looser). We may conjecture norms regulating them play an individuating role: what distinguishes one assertive from another is the characteristic normative constraint that it imposes on the speaker, in addition to TAR. If this is the case, jointly with TA, the norms of assertion thus provide the missing sufficient condition for identifying assertions – an alternative criterion to the definition I proposed in chapter III.

The view that assertives can be distinguished based on the constraints on their performance has already been defended in the literature. Searle’s influential proposal (Searle 1984, Searle and Vanderveken 1985) takes each family of speech acts (assertives, declaratives, etc.) to have a common direction of fit, and distinguishes illocutionary acts within each family based on the different rules to which they are subject. Turri (2010) presents the same ideas under a slightly different guise[[103]](#footnote-103). My suggestion to pair a truth-aim with non-factive norms is in line with this tradition. Its innovative feature is that it takes all assertives to be subject to TAR, so that any additional norm must be consistent with it.

To be sure, while I have mentioned that in assertion TAR could be paired with JR and BR, this is not the only way to achieve my proposal to pair a truth-aim view to a non-factive account, and I do not wish to commit to any particular solution in this respect. Even if the TAR-JR-BR trio seems to me the most reasonable candidate, there are several ways to enforce my proposal: any non-factive account paired with TA survives my proposed puzzle, and passes the test for the rule-aim distinction.

## 5. Objections, replies and clarifications

### 5.1 The source of normativity

Authors familiar with the literature on the norm of belief will perhaps find my explanation puzzling. Many authors have argued that beliefs aim at truth (*e.g.* Dummett 1959, Williams 1970, Platts 1979:257, Velleman 2000, Noordhof 2001, Steglich-Petersen 2006, 2008, see also Chan 2013), and it is certainly a standard view that beliefs have a *word-to-world* direction of fit. It is quite natural, then, to wonder whether my explanation could be constructed the other way around. It could be argued, in other words, that assertion aims at truth only insofar as beliefs aim at truth, if we understand assertions as the outward manifestation of a belief (cf. my discussion of ACF3 in III.1.2). On this view, the aim of truth is essential to beliefs rather than assertions, and only indirectly applies to the latter; similarly, norms like TAR regulate belief rather than assertion, and they impose normative constraints on assertion only indirectly, insofar as an assertion should express a correctly formed belief.

All these observations are fairly plausible, but they are not objections specific to my view. Authors like Bach (2007, 2008) and Hindriks (2007), for instance, have argued that KR is the norm of belief rather than assertion, and that assertions are constrained by KR only indirectly, in virtue of the fact that they express a belief. In other words, every account of the norm of assertion is vulnerable to an objection construed along these lines – for any *normative* account of assertion, we can construe a *descriptive* alternative that places normativity at the level of beliefs[[104]](#footnote-104). Similarly, for any *teleological* account of assertion (like mine), we can construct an alternative one that places the *aim* not at the level of assertion, but at the level of belief. This kind of challenge, then, is one to which virtually every account on the market is exposed.

More importantly, this dissertation is not really concerned about the source of normativity. The central desideratum of my account is to make the right predictions with respect to the permissibility of each and every assertion. Since this can be done without a story about the source of normativity (be it belief or assertion), the correctness of my account is independent of such a story. To be sure, in passing, I have hinted at a possible story about the source of these normative constraints (deriving it from a principle regulating cooperative action), but I have also stressed how nothing hinges on such story. If it can be proved that the source of normativity is better identified in a different way, such proof should be viewed as an improvement of the proposed account, rather an objection to it.

### 5.2 Primary and secondary violations

One fundamental assumption in this chapter was that factive accounts predict that unlucky assertions are impermissible. However, some factivists deny this prediction, and insist that unlucky assertions are permissible according to their view. This claim is controversial: if factivism requires truth for proper assertion, how can a false (even though unlucky) assertion be permissible?

The claim that unlucky assertions are deemed permissible by factive accounts relies on a distinction first introduced by Keith DeRose (2002:180, followed by Weiner 2005, Whiting 2012, Pritchard 2013:163-9, Pelling 2013, Benton 2013, Boyd 2014) who influentially argued that there are two ways of complying with a rule. *Primary compliance* tracks our ordinary understanding of what it means to comply with a rule: whenever you follow a rule R, you comply with R in the primary sense. *Secondary compliance* involves cases in which you reasonably believe that you are following the rule, but despite your belief you are actually violating the rule. Your infraction is ‘in good faith’, as you are unaware of the violation. Since you are trying to comply with the rule, there is a ‘secondary’ sense in which you are following it. To summarise DeRose’s distinction:

Primary compliance: A’s action *conforms* to R

Secondary compliance: A *believes* that his action *conforms* to R, but his action actually *violates* R

Primary violation: A’s action *violates* R

Secondary violation: A *believes* that his action *violates* R, but his action actually *conforms* to R

This distinction allows factivists to revise their predictions about unlucky assertions. Even if unlucky assertions involve a *primary violation* of factivist rules (TR/KR), they also involve *secondary compliance* to the rule. In the Unlucky Assertion fake barn example (cf. 2.2), Jacopo reasonably believes that he is following TR/KR (he reasonably believes that there is a yellow barn on his left), but he is actually violating TR/KR (there is no barn, just a barn facade). He is violating factive rules in the primary sense, but complying with them in the secondary sense[[105]](#footnote-105). This yields a new solution to the puzzle of reconciling Permissible Falsity and Falsity-Criticism. According to factive accounts, unlucky assertions are both, in one sense, appropriate (as they are in *secondary compliance* with the rule, satisfying Permissible Falsity) and, in another sense, inappropriate (they are nonetheless in *primary violation* of the rule, satisfying Falsity Criticism). This seems in line with our contrasting intuitions about unlucky assertions: they appear to be both permissible and, in some other sense, wrong.

However, it is not clear that replies based on the primary/secondary distinction are viable for the factivist. Against DeRose and his followers, several philosophers have rejected the primary/secondary distinction. A growing number of scholars now consider it to be a spurious distinction (Douven 2006:478-480, Lackey 2007:604, Stone 2007:100, Engel 2008, Koethe 2009:631fn16, Cappelen 2011:46, Greenough 2011: fn29, van der Schaar 2001:195-6, Kvanvig 2011: 242, Hinchman 2013:641fn6, Pagin 2016:17-22). In what follows, I will briefly summarise some objections to the distinction, to lend support to the view that this is not a genuine distinction between two ways of violating or following a rule, but rather a *post-hoc* expedient to make unlucky assertions fit into the factivist framework.

A first criticism comes from Lackey (2007). Broadly, her argument is that the distinction is spurious because it cannot be extended to any other known practice: in any given practice or game with which we are familiar, secondary compliances are simply not compliances (they are violations), and secondary violations are simply not violations (they are compliances)[[106]](#footnote-106). To see this, consider the following example by Turri and Blouw (2015, simplified here).

Chess Tournament

Nicolas just accepted an invitation to participate in a chess tournament. The invitation states that in the tournament each player is allowed to take a maximum of 5 minutes to make a move. But Nicolas’s fierce competitor, Vladimir, has intercepted Nicolas’s invitation and modified it so that it states that the maximum time to make a move is 10 minutes. Nicolas isn’t aware that Vladimir did this; as a result, during the first match, Nicolas takes 7 minutes to move the rook, and gets disqualified.

In the example, Nicolas is in *primary violation* and *secondary compliance* with the rules of the tournament (like Jacopo’s unlucky assertion). However, we have the intuition that even if Nicholas reasonably believes that he followed the rule, there is no ‘secondary’ sense in which he has, as a matter of fact, followed the rule. Lackey believes that this intuition is not restricted to special cases, but generalises to any rule-governed activity. There are no games or practices to which the primary/secondary distinction applies meaningfully: if you reasonably but falsely believe that you did not violate a rule, you still have violated that rule; you surely have *tried* to followed it, but you have not as a matter of fact followed it (Turri and Blouw 2015).

Pagin (2015) shares Lackey’s worry, but focuses on a methodological criticism: he points out that the distinction can be exploited to rebut almost any counterexample to epistemic accounts of the norm of assertion.Consider another example: the truth-norm (TR) incorrectly predicts that *lucky* assertions are permissible. But this counterexample can be dismissed by appealing to the primary/secondary distinction (cf. Weiner 2005). A lucky assertor is *secondarily* violating TR (he reasonably believes he is violating it), so that there is a secondary sense in which TR deem his action impermissible. Pagin worries that the same strategy could be indefinitely generalised:

If we consider the matter abstractly, we can imagine four mutually incompatible theories, A, B, C, and D, and a set of intuitive judgments. Suppose that 25% of these judgments are primarily proper with respect to theory A, while the remaining 75% are primarily improper but secondarily proper. Suppose that the other three theories are in the corresponding situation. By appeal to the primary/ secondary distinction together with corresponding differences between the theories as to what properties the intuitions are tracking, we get the following result. For each of the theories, a defender can claim that his theory is supported by all intuitive judgments, for he can explain 25% of the intuitions as tracking primary correctness and the remaining 75% as tracking secondary correctness. He explains those which track secondary correctness because he derives the secondary norm from his primary norm, together with some additional general premise. (Pagin 2016:20)

In sum, Pagin believes that if we accept the primary/secondary distinction, each account of the norm of assertion is radically underdetermined: there is no way to find one piece of empirical evidence that supports one account over the other. Since the distinction blurs the differences between each account’s predictive content, thereby making it impossible to test them against either intuitions or empirical data, the distinction is highly undesirable.

We might accept these criticisms, and Lackey’s plausible objection that the secondary violations are still violations (and secondary compliances still compliances). This is not to say that the distinction is arbitrary. On the contrary, it captures the difference between the violations/compliances that are intentional and those that are not. The distinction is meaningful, because it is relevant to *evaluative judgments*: *ceteris paribus,* intentional (primary) violations are *blameworthy,* and unintentional (secondary) violations are *blameless*. Starting from these observations, the factivist can reformulate his argument while avoiding the primary/secondary distinction. Even if unlucky assertors violate factive rules, they do so in a blameless fashion. This is enough to explain the Permissible Falsity Linguistic Behaviour within a factivist framework: we do not criticise unlucky assertors for saying something false because their violation of the relevant factive norm is blameless. As for the Permissible Falsity Linguistic Intuition (the intuition that unlucky assertions are permissible despite being false), the factivist can simply deny it, or dismiss it as being distorted by the fact that we judge such assertions to be *overall* blameless (Turri and Blouw 2015).

A first difficulty with the revised argument is that while it seems to address Lackey’s objection, it does not address Pagin’s. But let us grant, for the sake of the argument, that Pagin’s objection is unfounded. This would lead us to a stalemate: KR and JR are equally supported by the linguistic data. To decide in favour of one account over the other, we need to bring in some independent data. Luckily, we already have the resources to get out of this impasse. In 4.1, I have introduced the Test for the Rule-Aim Distinction. The test showed that, *all things the same*, truth is better understood as the aim, rather than the rule, of assertion. A key intuition behind the test was that rules do not typically allow for inadvertent (and therefore *blameless*)violations by careful and competent agents*.* This intuition was supported by the observation that a factive rule of assertion would not be *fair*, as it would be in violation of the ‘ought ought to imply can principle’. This suggest that, *ceteris paribus,* the account predicting that we cannot typically have *blameless* violations of the rules should be preferred over the one that typically allows for such violations.

To be sure, the test does not predict that any rule that allows for blameless violations cannot be a fair (or genuine) rule – the test only target (putative) rules that systematically allow for this sort of violation. Consider again the Chess Tournament scenario. Here the ‘5-minutes rule’ allows Nicholas to commit a *blameless violation,* but there is an important sense in which the scenario is atypical: hardly any chess player can say to have experienced a comparable situation. By contrast, I have already argued (cf. 2.2) that there is an important sense in which *unlucky assertions* are not atypical cases: every speaker is familiar with asserting what they reasonably (but wrongly) believed to be true.

A further difficulty for the factivists is that appealing to the blamelessnessof inadvertent violations is perfectly sensible when we consider non-epistemic rules, but less plausible when we consider epistemic ones. Epistemic rules dictate which *epistemic position* is required for a permissible action, whereas non-epistemic rules deem actions permissible or impermissible independently of the agent’s epistemic position. In the context of non-epistemic rules (like the rules of chess, or traffic regulations) it makes sense to *excuse* an agent for *reasonably believing* that he did not violate a rule. But in the context of epistemic rules, the very point of the rule is to establish whether *reasonable belief in the truth of a proposition* is sufficient or not (and necessary or not) for permissible action. If the point of the rule is setting a specific epistemic standard, it is not clear on which grounds we can excuse an agent for failing to meet that very epistemic standard. In other words, if the *point* of choosing KR over JR is that the former requires assertors to meet an epistemic standard higher than *reasonable belief*, it makes no sense to excuse assertors for failing to meet any epistemic standard higher than reasonable belief. All in all, it seems that the argument from blameless violations cannot save the factivist cause.

### 5.3 Challenges

The truth-aim account correctly predicts that I can criticise someone for asserting something false. For instance, I can reply to a false assertion: “Hey, wait a second! That is not true!”. However, this kind of challenge has the form of a criticism for failing to follow a *rule*, rather than failing to meet an *aim*. For instance, in playing chess with a friend I would not say “Hey, wait a second! That is not a good move!” when he fails to make a move that would lead him to meet his aim to checkmate. By contrast, I would say “Hey, wait a second! That move is not allowed!” when my friend puts the bishop on top of the horse and moves the resulting piece along an L-shaped diagonal. It seems that we can challenge a move in a game when it violates a rule, rather than when it fails to meet an aim. If this is correct, then it is not clear why in the assertion-game we would be entitled to challenge moves that fail to meet the aim of assertion, *i.e.* assertions that fail to be true.

This objection, however, is based on a misconstrued analogy. Chess is a *competitive* game, and it is in the interest of adversaries that the other party does not make the best moves. For this reason, criticisms often fail to be voiced – but are still appropriate when they are produced by third party, *e.g.* by spectators. Assertion, as I construed it, is rather a practice governed by a presumption of *cooperation*. And in cooperative practices, it is perfectly appropriate to correct the other cooperating party for failing to meet a given aim. For instance, if I am teaming with Niccolò in playing Scrabble, before he lays down the letters I can stop him and tell him: “Hey, wait a second! That is not a good move!”, and suggest a better combination of words. This shows that a move in a game can be appropriately challenged both when it violates a rule and when it fails to meet an aim, so that the objection misses the point.

### 5.4 Knowledge as the aim?

The knowledge-rule account of assertion (KR) is certainly the most influential account on the market, with an impressive record of scholarship both in favour and against it. Whether knowledge is the rule of assertion not, it is certainly true that many authors believe that there is *some* link between assertion and knowledge (knowledge-first epistemologists first in line). Now, some of these authors may be tempted to suggest that, if knowledge cannot be the rule of assertion (for the reasons I presented), perhaps a link between assertion and knowledge can still be maintained by positing knowledge as the aim of assertion.

This strategy, however, is not easily pursuable. In the way I have defined rules and aims, they are mutually exclusive concepts. Proponents of the KR support their view by observing that the rule is able to explain the linguistic *impermissibility* of a certain set of actions. Understood as an aim, knowledge cannot explain these very actions as impermissible, but only as *unsuccessful*. The resulting account would therefore make wildly different predictions with respect to what is permissible to assert.

Suppose that proponents of the knowledge-aim account are ready to bite this bullet, and to endorse the revised predictions. In this case, they clearly have the option of supplementing their view by pairing it with the ‘knowledge-aim rule’ KAR – pretty much in the same way that I suggested to pair TAR with the truth-aim view.

(KAR) You should: assert *p* only if you are thereby trying to assert a proposition you know

I believe that such an account would be a plausible proposal. It would clearly be a non-factive account, making predictions very distant from KR ­– but nonetheless analogous to it, insofar as it still enforces a link between assertion and knowledge. Its advantage over my proposal to pair the truth-aim account with non-factive rules (as TAR/JR/BR) would be its simplicity. Its disadvantage would be that it would fail to generalise to other illocutionary acts: other assertive speech acts, like guessing, do not aim at knowledge in the envisaged sense. A further problem is that it could not inherit my proposed understanding of ‘aiming’ in terms of direction of fit: it seems implausible to postulate a *word-to-knowledge* direction of fit. More generally, it seems difficult to spell out what aiming at knowledge could mean. Trying to determine whether these challenges can be met goes beyond the scope of this thesis. For the present purposes, it is sufficient to note that if a knowledge-aim account can be plausibly defended, it would need to address these difficulties.

### 5.5 Alleged asymmetries

I have construed the rule-aim difference by relying on an analogy with games, claiming that for a practice to aim at X is to have X as a success condition. However, while aim-directed moves in a game (such a scoring a penalty) can be ordinarily described as successful or unsuccessful, assertions do not fit this category of assessment. Contrarily to what I have claimed, we do not ordinarily praise true assertions *as successful*, and we do not criticise false ones *as unsuccessful*.

This objection fails for two reasons: it relies on disputable observations and on a misconstrued analogy. It relies on disputable observations because it is not clear that, in ordinary language, we can *always* describe aim-directed moves in a game as successful and unsuccessful. For instance, my picking the most valuable combination of words of my Scrabble hand and laying it on the table certainly meets the purported aim of my turn, but my laying down that combination would ordinarily be described as a good or excellent move, rather than as a successful one.

Furthermore, the proposed analogy is misconstrued because ‘succeeded in asserting’ is different from ‘succeeded in *scoring* a penalty’, in that the latter expression refers to the success condition of the action (score), while the former to the action itself. The appropriate comparison in football be with ‘succeeded in *shooting* a penalty’, which is not an expression of praise in football. Alternatively, the appropriate comparison in assertion would be ‘succeeded in asserting the *truth*’ or ‘succeeded in describing an *actual state of affairs*’, and these are appropriate ways to praise of an assertion. Once again, ordinary language judgements must be taken with a pinch of salt.

A related problem concerns a putative asymmetry between rules and aims. I have claimed that aims, like rules, can ground both praises and criticisms of a given action. However, while we generally praise players for meeting an aim (*e.g.* scoring a goal), we rarely praise players for following a rule (*e.g.* it is inappropriate to praise a competent player who moves the bishop correctly). Since we do not typically praise true assertions for being true, our behaviour is thus in line with a putative truth-rule of assertion, rather than a truth-aim.

This objection relies on an unwarranted generalisation, according to which, since it is often appropriate to praise players for succeeding in scoring a goal, the same must apply to other aims. Appropriateness of praising success seems rather to be a *function of the difficulty of meeting the aim*: the less difficult to meet the aim, the less appropriate to praise success. This observation correctly predicts that compliance with a rule is not a good ground for praise, as typically one cannot try but fail to follow a rule (cf. 4.2). As for aims, the prediction is that meeting them can be a good ground for praise (as typically one can try but fail to meet an aim); however, such appropriateness comes in degrees, and the easier success is, the less appropriate praising is.

If this is correct, we should be able to identify aims that are very easy to meet and that it is not really appropriate to praise. This is indeed the case. Consider football again: we do not generally praise football players for succeeding to complete a passage, because completing a passage is a relatively easy task. However, consistently with the proposed principle, the more difficult the passage (*i.e.* the higher the possibility to fail), the more appropriate the praising becomes.

Similar observations apply to assertions. We do not generally praise true assertions, because in the majority of cases asserting the truth is easy: I cannot easily try and fail to tell you truly that my finger hurts, or that I put two spoonsful of sugar in your coffee. However, when the possibility of failure becomes relevantly high, praise is appropriate. For instance, we do praise Galileo for his strenuous insistence in asserting that Earth turns around the Sun. Similarly, in a context of disagreement (*e.g.* on whether Canberra is the capital of Australia), it is appropriate for the wrong party to praise the right party’s true assertion when the matter is settled.

The bottom line is that the inappropriateness of praising most assertions seems to be directly proportional to the relative simplicity of succeeding in telling the truth, rather than a function of truth being the rule of assertion. More importantly, since *some* assertions can be praised, the objection can be turned upside down. It could be argued that it is far more difficult to explain praising in terms of compliance with rules. If rules compliance does not indeed elicit this sort of praise, the praise of these assertions will turn to be a datum in our favour of the truth-aim account, rather than of the factive ones; otherwise, the datum simply does not bear on the issue.

## 6. Conclusions

This concludes the second and last part of the dissertation, discussing the norms regulating assertions. In these two chapters, I have put pressure on a number of ideas that are currently very popular in the philosophical literature. In chapter V, I have questioned the hypothesis that the C-rule is *constitutive* of assertion, arguing that there is no significant way to interpret this claim. This provided new grounds to doubt some further disputable assumptions in the debate: that there is *only one norm* regulating assertion, and that assertion can be *defined* as the unique speech act that is uniquely subject to the norm. While chapter V assessed the solidity of the hypothesis that assertion is governed by a constitutive norm, chapter VI tried to determine which norm constitutes the most plausible candidate. More specifically, I have dealt with the debate between factivist and non-factivist accounts, and shown that non-factivist accounts (once paired with the view that assertions aim at truth) are preferable to factivist ones.

Together with part A, this dissertation offers a general model of what it is to claim that something is true (*i.e.* to assert), and of the obligations that come with such claims. In the chapter on lying, I have focused on definitional issues. I have defined assertion as a speech act in which the speaker *presents a proposition as true*, thereby *committing himself to the truth of the proposition*. I have shown that one can assert directly, by plainly stating something, or indirectly, by performing some other speech act that has assertoric force – such as promising, assuring or admitting that something is the case. Part B has completed this picture. I argued that in *presenting a proposition as true*, an assertor determines a *success-condition*: the assertion is successful in representing reality only if the proposition is true. In this sense, we can say that assertions *aim at truth*, and that they are correct when true and incorrect when false. Second, I presented a general picture of the normative constraints that apply to assertion. For the purpose of defining ‘lying’, I had only considered responsibilities that come after one asserts (*downstream* responsibilities) – what I called *being* *committed to the truth of the proposition*. In the second part of the dissertation, I considered the responsibilities that come before one asserts (*upstream* responsibilities); namely, that one should ensure that every asserted proposition meets the conditions dictated by the norms of assertion. Some questions remain open; for instance, I have not determined whether assertion is subject to further, stricter norms, such as a norm requiring assertors to only assert what they *rationally* believe. These are questions for other works to investigate; hopefully, the account of assertion that developed in this thesis has provided a novel and solid theoretical framework to address them.

# VII. Conclusions

That Lye, shall lie so heavy on my Sword,  
That it shall render Vengeance, and  
Revenge, Till thou the Lye-giver, and that  
Lye, do lye In earth as quiet,  
as thy Fathers Scull.

Shakespeare, *Richard II*, 4, 1

In this dissertation, I have considered two important ways in which assertions (claims that something is true) can fail to meet basic conversational expectations. The first part (part A) has dealt with lying, that is, assertions that are *insincere*. Most of my attention has been devoted to identifying a definition of lying that reflects our use of the term. In developing a definition, I have also provided a description of what lying is, shedding new light onto this phenomenon and related ones, such as insincerity and deception. Furthermore, I have explored different ways in which a speaker can lie: with or without an intention to deceive, more or less insincerely, asserting directly or via another speech act.

In the second part (part B), I have considered the influential view that assertions are subject to a single epistemic norm. I criticised some assumptions that philosophers make in this debate – namely the assumption that the norm is constitutive, individuating and unique of assertion. I have then turned to the question of whether assertions are permissible only if they are true. I have defended an original solution: that assertions (purport to) aim at truth, but can sometimes be permissible even if they are false. On my view, speakers are supposed to attempt to only say what is true. If they try their best to tell the truth but fail, their assertion is incorrect, but not impermissible.

Throughout the thesis, I have introduced different views, surveyed counterexamples, presented several definitions, and drawn a number of conclusions – some of which might have been lost in a sea of words. Admittedly, it could have been sometimes difficult for readers to keep track of the red wire connecting the different issues covered in this dissertation. For this reason, I have resolved to present my conclusions in a very schematic form, sacrificing discursiveness to conciseness and clarity. In what follows, I summarise each chapter in a few bullet points. The main conclusions established in each chapter are highlighted in bold. Hopefully, this will give the reader a bird-eye view of the whole dissertation, helping them to visualise my key points of contention and understand how the main themes of the thesis are connected.

Part A: Lying

What is lying? Philosophers have long attempted to find an answer to this elusive question. They generally agree that you lie only if you intentionally say what you believe to be false, but they disagree about which further conditions can provide a sufficient analysis of the concept of lying. The first part of the thesis has attempted to provide a solution to this longstanding debate.

Chapter 2: Deceptionism vs Nondeceptionism

Deceptionist philosophers argue that intending to deceive is a necessary condition for lying. They defend, in some version or another, the *intention to deceive condition* (IDC) as a necessary condition for lying. Most deceptionist philosophers also argue that the IDC, jointly with the *statement condition* (the speaker must state that *p*) and the *insincerity condition* (the speaker must believe that not *p*), offers a sufficient analysis of the concept of lying. In chapter 1, I have considered whether deceptionist philosophers are right, and established the following conclusions:

* **Specific versions of the IDC are subject to counterexamples:** A review of the literature shows that no matter how the IDC is phrased, there are counterexamples to each version of the IDC.
* **The deceptionist analysis is too broad**: It is possible to produce counterexamples to the sufficiency claim made by deceptionists, namely cases of non-assertoric deception – fictional or ironic utterances that are meant to deceive. The fact that these utterances are intuitively not lies shows that the deceptionist analysis of lying is too broad, unless it is supplemented with a further condition, requiring that the speaker asserts what he is saying.
* **The deceptionist analysis is too narrow:** *Bald-faced lies* are insincere statements uttered without any deceptive intent. Intuitively they are lies, but the IDC (no matter how it is phrased) is unable to count them as lies, as they are not intended to deceive. Counterexamples based on bald-faced lies disqualify the IDC as a necessary condition for lying.
* **Deceptionist replies to objections based on bald-faced lies are misguided:** Some deceptionist authors have challenged counterexamples based on bald-faced lies. Simply put, they have replied either that bald-faced lies are indeed intended to deceive, or that they are not assertions and therefore not lies. In response to these arguments, I have shown that:
  + **Most bald-faced lies are not intended to deceive**
  + **Some bald-faced lies are indeed assertions**

If my argumentations are right, bald-faced lies still stand as a counterexample to the IDC, so that intending to deceive is not a necessary condition for lying.

Taken altogether, these observations make a strong case against deceptionism. The main conclusion of chapter 1 is that **intending to deceive is not a necessary condition for lying**; on top of this, several observations point out to the further conclusion that **genuinely asserting a proposition is a necessary condition for lying** – a hypothesis tested more extensively in Chapter 2.

Chapter 3: Lying and Asserting

Most non-deceptionist philosophers agree that lying can be defined as an insincere assertion. The bone of contention between alternative non-deceptionist definitions is how to define assertion. In Chapter 3, I have attempted to develop an assertion-based non-deceptionist definition of lying, and concluded that:

* **Every known assertion-based definition is subject to counterexamples:**
  + Carson’s definition of assertion as *intending to warrant* cannot account for proviso-lies.
  + Fallis’ definition of assertion as *a statement uttered in a context in which the speaker believes the Gricean Maxim of Truthfulness to be in effect* incorrectly counts ironyas lying.
  + Fallis’ first alternative definition, that implicitly takes assertion to require *intending to communicate what you say,* is underdetermined, and too broad if interpreted in a literal way.
  + Fallis’ second alternative definition as a *representing yourself as believing the stated proposition* is more promising, but it is also too broad
  + Stokke’s definition of assertion as *proposing the stated proposition to become common ground* falls victim of a known counterexample by Fallis, of which I vindicate the effectiveness.
* **No known definition can deal with my three ‘speech-act theoretic puzzles’:**

Taking inspiration from speech-act theory, I identified three desiderata for a definition of lying that traditional accounts have failed to meet so far. These desiderata are:

* 1. Counting some performative utterances as lies
  2. Identifying which performative utterances can be lies and which cannot
  3. Allowing for attitudes other than belief (namely intentions) to determine whether an assertion is insincere or not
* **Speech-act theory provides the theoretical tools to meet these challenges:**

In the light of the failure of both deceptionist and non-deceptionist account*s* to solve the speech-act theoretic puzzles, I resorted to speech act theory to develop an alternative definition. The result is the following (non-deceptionist, assertion-based) definition of lying:

**In uttering U, S lies to A about *p* iff:**

1. *S thereby asserts that p, i.e.:*
   1. *S expresses p*
   2. *S presents p as an actual state of affairs*
   3. *S takes responsibility for p being an actual state of affairs*
2. *S is insincere*

Condition 1 allows for assertions to be performed either directly or by means of **illocutionary entailment**, a notion that I borrow from Searle and Vanderveken:

*The performance of an illocutionary act F1(p)* ***entails*** *the performance of another illocutionary act F2(p) iff in the context of the utterance it is not possible for S to perform F1(p) without performing F2(p) – so that if S performs F1(p), S also performs F2(p)*

The main implication of this view is that you can assert something by performing another illocutionary act; for instance, by saying “I promise that p”. The notion of illocutionary entailment provides a clear criterion to distinguish between the illocutionary acts that can be used to assert (and therefore to lie) and the ones that cannot.

This provides a solution for the three speech-act theoretic puzzles:

1. Explicit performatives can be counted as lies, because this definition identifies the content of the utterance in a non-descriptivist fashion – as prescribed by condition (1a) in the definition.
2. Condition (1b) and (1c), paired with the notion of illocutionary entailment, **identify** **which illocutionary acts can count as lies**: **assertives** (except those weaker than assertion)and **commissives**.
3. The proposed definition allows for attitudes other than beliefs (e.g. intentions) to determine whether an utterance is insincere. A more detailed discussion of insincerity is developed in chapter 4.

Chapter 4: Insincerity

In this chapter, I review some problems concerning the modelling of the notion of insincerity, especially in the light of the third speech-act theoretic problem highlighted in Chapter 3 (how to extend the insincerity condition to attitudes other than belief?).

* **Speech act theory provides a** **simple definition of sincerity and insincerity** that can apply to attitudes other than belief:
* *The performance of an illocutionary act F(p) that expresses the psychological state Ψ(p) is* ***sincere*** *IFF in uttering F(p), S is in Ψ(p)*
* *The performance of an illocutionary act F(p) that expresses the psychological state Ψ(p) is* ***insincere*** *IFF in uttering F(p), S is not in Ψ(p)*

If we apply this model to assertion, we have that since assertions express beliefs, an assertion is sincere iff the speaker believes that *p,* and insincere iff the speaker does not believe *p.*

This model, however, is incomplete, as it does not provide clear predictions for cases in which one illocutionary act entails another. To remedy this gap. I proposed the following definition of insincerity in cases of entailment:

* **Definition of insincerity by entailment**
* *When the performance of an act F1(p) expressing the psychological state Ψ1 (p)* ***entails*** *the performance of another illocutionary act F2(p) expressing the psychological state Ψ2(p), the utterance is insincere iff the speaker either is not in Ψ1 (p), or the speaker is not in Ψ2(p), or both*

To test this general model, I have derived from this definition the insincerity conditions for promising. A promise that *p* expresses an intention to *p,* but it also illocutionarily entails an assertion that *p.* Consequently, its insincerity conditions are:

* **Definition of insincere promise (lying by promising)**
* *A promise that p is insincere iff the speaker either does not believe that p, or does not intend to p, or both*

When this insincerity condition obtains (i.e. the promise is insincere), since a promise entails an assertion (cf. Chapter 3), the speaker is also lying.

* **This definition of insincerity reflects laypeople’s intuitions about lying**

Most philosophers agree that a definition of lying should reflect the intuitions of ordinary speakers, but few test their theories against such intuitions. I tested my view (that speech acts other than assertion can be lies, under the conditions identified by the definition of insincerity by entailment) against the judgment of ordinary speakers. My experimental study demonstrates that **only my account makes the right predictions** when it comes to promises involving insincere intentions but sincere beliefs, or insincere beliefs but sincere intentions**.**

Given the success of my account in modelling illocutionary acts other than assertions, I moved on to consider a further philosophical problem concerning insincerity. Most philosophers writing on lying accept a ‘dichotomic view’ of insincerity:

* **Dichotomic view:** A statement is sincere when the speaker believes it to be true, and insincere when the speaker believes it to be false, *tertium non datur.*

In order to test this view, I introduced two kinds of lies that cause troubles for this view:

* **Fuzzy lies:** the speaker has a belief that assigns a graded truth value to a proposition, but his assertion assigns a different graded truth value to that proposition.
* **Graded-belief lies**: thespeaker has a graded belief about a proposition, but his assertion expresses a different degree of belief in that proposition.
* **The dichotomic view has troubles dealing with the fuzzy lies and graded-belief lies:**

Both fuzzy lies and graded-belief lies involve a graded difference between *what is believed* and *what is asserted*. The problem for the dichotomic view is that this difference that cannot be cashed out in terms of a simple opposition between believing and not-believing.**These cases call for a more fine-grained criterion** to determine which degree of discrepancy between what is asserted and what is believed is needed for an assertion to count as insincere, and therefore as a lie.

After reviewing some alternative accounts of insincerity, I proposed the following one:

* **Comparative insincerity condition:**

*In asserting p, S is insincere iff S is more confident in ¬p than S is confident in p*

Some problems, however, remain open. A theory of lying should still explain the following data:

* The degree of discrepancy between what is asserted and what is believed being the same*, overstating one’s degree belief is worse than understating. The former generally counts as lying, while the latter generally does not.*

To explain this, I offered the following explanation:

* **Intensity modifiers affect both the degree of belief expressed by the utterance and its illocutionary force**.
* **Mitigating modifiers can decrease the illocutionary force to such an extent that the utterance no longer counts as an assertion, and therefore as a lie.**

A second issue is that this account of ‘graded insincerity’ only applies to beliefs. In the concluding section of the chapter, I integrate this model with my previous observations about how insincerity applies to attitudes other than belief, to recover a general definition of lying:

* **Definition of lying by entailed assertion**

*In successfully uttering an illocutionary act with content p that expresses an attitude* *Ψ(p) and entails an assertion with content p, S lies to A about p iff:*

1. S thereby asserts that *p,* i.e. (a), (b), and (c) obtain
2. Either S is more confident in *p* that S is confident in ¬p, or S is in Ψ(¬p) more than S is in Ψ(p), or both

In the case in which the speaker is asserting directly, this definition will reduce to the following:

* **Definition of lying by asserting**

*In successfully uttering an illocutionary act with content p*, *S lies to A about p iff:*

1. S thereby asserts that *p* (i.e. (a), (b), and (c) obtain)
2. S is more confident in *p* than S is confident in ¬*p*

Part B: Assertion and its norms

Assertions are defective when they are insincere. But this is not the only way in which they can be defective. In other words, it seems that you cannot assert just about anything, as long as you believe it. It seems that you need some grounds for your assertion to be an acceptable conversational move. To account for this intuition, philosophers have hypothesised that assertion is subject to a norm that requires speakers to assert only when a given epistemic standard is met. More specifically, many philosophers accept Williamson’s (2000:241) hypothesis:

**Williamson’s hypothesis**  
*What are the rules of assertion? An attractively simple suggestion is this. There  
is just one [constitutive] rule. Where C is a property of propositions, the rule  
says:  
(The C-rule) • One must: assert p only if p has C.*

In this part of the thesis, I considered the extent to which Williamson’s hypothesis is a plausible conjecture, and then introduced my own account of the norms and aims of assertion.

Chapter 5: The norm of assertion: a constitutive norm?

In Chapter 5, I took issue with the claim that the norm is constitutive. I argued that:

* **Williamson’s definition of ‘constitutive rule’ departs significantly from the orthodoxy**, as shown in the rows 2-3 of the following table:

|  |  |  |
| --- | --- | --- |
| **#** | **Constitutive rules – orthodox view** | **Constitutive rules – Williamson** |
| 1 | Establish (i.e. define) the practices they regulate | Are essential to (i.e. define) the practices they regulate |
| 2 | Cannot be ‘properly’ violated | Can be violated |
| 3 | Typically take the form of ‘count-as’ locutions | Typically take the form of imperatives |

* **Some authors interpret the notion of ‘constitutive rule’ in the traditional sense, others in Williamson’s sense.**
* **Both interpretations of the hypothesis are problematic**:
  + The traditional interpretation is incompatible with the hypothesis.
  + The alternative interpretation renders the hypothesis that the norm is constitutive uninteresting, as all rules possess the properties identified as distinctive of constitutive rules.
* In conclusion, **philosophers are better off abandoning the claim that there is a norm that is *constitutive* of assertion**.
* Once the claim is abandoned, **also the claim that the norm is unique and the claim that the norm identifies assertion** (assertion can be defined as the only speech act that is uniquely subject to the norm) **lose their appeal**.

Chapter 6: Assertion and Truth

Can a false proposition ever be assertable? Depending on the response to this question, accounts of the norm of assertion can be divided into two families:

* ***Factive accounts***: only true propositions are assertable.
* ***Non-factive accounts***: some propositions are assertable even if they are not true (granted that they meet some epistemic standard).

These accounts are motivated by different intuitions:

* ***Falsity-Criticism***(supporting factive views): true assertions are intuitively correct, and their truth is apt to be praised; false assertions are intuitively incorrect, and their falsity is liable to criticism.
* ***Lucky and unlucky assertions***(supporting non-factive views): some assertions are intuitively impermissible even if they are true; some assertions are intuitively permissible even if they are false.
* **Factive norms are too demanding**: factive norms seem to predict that there is no permissible course of action for a speaker in every case in which a speaker *S is asked whether p is the case, and S holds a false but justified belief about p.*
* **One cannot always derive impermissibility from incorrectness**. The terms ‘correct’ and ‘incorrect’ are ambiguous between two meanings. One is alethic (‘correct’ means ‘true’), another is normative (‘correct’ means ‘permissible’). The *argument from correctness,* that derives factive norms from the *Falsity-Criticism* intuition, relies on an equivocation between these two readings. The conclusion is that **intuitions about the incorrectness of false assertions** (which constitute an important part of the Falsity-Criticism intuition) **do not bear on the issue** of whether permissible assertion requires truth.
* **The Falsity-Criticism intuition can be explained within a non-factivist framework:** even if one cannot derive the impermissibility of false assertion from their incorrectness, it is still true that false assertions are criticisable, and an account of assertion needs to explain why. The factivist explanation is not the only option: **criticisms about the falsity of false assertions can be explained in terms of failure to meet the aim, rather than the norm, of assertion**.
* **An intuitive test suggests that assertion is the aim, rather than the rule, of assertion:**

**Test for the truth-aim distinction**

Where A is an activity, C a condition applying to A, and S an agent,the truth of (T) is *prima facie* evidence that C is the aim of A, whereas the falsity of (T) is *prima facie* evidence that C is the rule of A:

(T) S can typically try but fail to meet C.

* **If truth is the aim of assertion,**
  + **truth cannot be the rule of assertion, so that factivist views should be abandoned in favour of non-factivist ones**.
  + **assertion is arguably subject to a non-factive rule that requires the speaker** to try to conform to the purported aim of its action, *i.e.* **to try** **to tell the truth**.
  + **assertion is arguably subject to further and stricter rules** that differentiates assertion from other assertive speech acts, such as hypothesising or testifying**. Rules like BR and JR are plausible candidates:**
    - **BR**:Assert that *p* only if you believe that *p.*
    - **JR**: Assert that *p* only if you reasonably believe that *p.*

This dissertation has provided a systematic characterisation of two kinds of communicative faults: lies and improper assertions. The first part has developed fine-grained definitions for the concepts of *lying*, *asserting* and *insincerity*; the second part an account of what makes an assertion permissible and successful *qua assertion*. Overall, this dissertation offers a simple but comprehensive account of the communicative vices that can affect any claims to the truth of a proposition.

# Acknowledgements

I would like begin my acknowledgments by expressing my *sincere* gratitude to three fantastic supervisors, to whom I am hugely indebted. Jennifer Saul patiently read all the chapters of this thesis, some of which several times, always providing helpful comments and enthusiastic support. She was for me a patient and supportive guide, from the day I first moved to the UK to the final steps of my journey as a PhD student. I am extremely grateful for her tetratricotomic work of correction, and for the many lies that she allowed me to tell her and that she told me – ranging from the infamous affair between Cameron and his pig, to the NASA conspiracy to send slave children to Mars. Paul Faulkner also had a crucial role in the redaction of this dissertation. He read several portions of it, offering illuminating (and often challenging) feedback. In the event that my dissertation is a valuable contribution to the scholarship, it is to an important extent his merit. During the pleasant and instructive semester that I spent at the LOGOS group in Barcelona, Manuel Garcia-Carpintero volunteered to offer his supervision, always providing encouraging support, and precious comments to my work.

As well as my supervisors, I am thankful to several graduate students that I met during my PhD. Two friends in particular have been invaluable for my thesis. Miklos *Khurty* has helped me to run my experimental studies (section 3 of chapter IV); without his support, that part of my thesis would have never seen light. Our collaboration eventually brought us to Germany, on an adventurous academic trip to present our research. I am grateful to Miklos for being a fantastic friend; not only that, but also for smoking all my cigarettes, folding my pyjamas, and teaching me how to play with SPSS and how to be a descriptivist. Ivan Milic has read everything I wrote about assertion, providing priceless feedback. Together, we launched a reading group on assertion at the University of Barcelona, during the sessions of which I developed and refined the key positions that I defend in the second part of my thesis. With Ivan I witnessed several dawns in Barcelona, and danced under the notes of Frank Sinatra: he was a precious colleague and a dear friend.

Several other friendships with graduate students shaped my PhD experience. I had the great fortune of always being surrounded by an extremely friendly and stimulating group of people, both at the University of Sheffield and at the University of Barcelona. To list everyone would be impossible. Nonetheless, I would like to mention a few. At the University of Sheffield, I would like to thank to Francesco Antilici, Graham Bex-Priestley, Josh Black, Stephen Bolton, Lewis Brooks, Corin Bruce, Adriana Clavel Vazquez, Alexandre Duval, Anton Eriksson, Ahmad Fattah, Damiano La Manna, James Lewis, Gonzalo Núñez Erices, Neil Williams, and the ‘Attic Crew’ as a whole. At the University of Barcelona, I would like to thank Toni Alexandrov, Delia Belleri, Samuele Chilovi, Roger Delafeu, Jojo Findl, Francesco Gallina, Narsi Marsi, Alessio Santelli, Carlos Muñoz-Suárez – and, obviously, Paco Murcia. Margherita Isella from the University of Milan deserves a special mention – we were never colleagues, but we met at more than one conference and had long, regular discussions on Skype, during which we both learned a great deal about insincere promises and gullible children.

I am also heavily indebted to several professors: some have read and commented parts of this thesis, others have provided suggestions or ideas during personal conversations or Q&As, others have provided other forms of support for its redaction. Again, it is a challenge to list them all, but it is worth trying. I am grateful to Franca D’Agostini (University of Milan - Statale), Carla Bianchi (University of Milan – S.Raffaele), Carla Bazzanella (University of Turin), Marta Dynel (University of Łódź ), Mark Jary (University of Roehampton), Rosanna Keefe (University of Sheffield), Paolo Labinaz (University of Trieste), Berislav Marušić (Brandeis University), Jörg Meibauer (University of Meinz), Sven Rosenkranz (University of Barcelona), Marina Sbisà (University of Trieste), Andreas Stokke (Uppsala University), Giuliano Torrengo (University of Milan - Statale), Alberto Voltolini (University of Turin), Alex Wiegmann (University of Goettingen), and Igor Zagar (Educational Research Institute).

This work has been presented at several conferences and workshops in Europe – more specifically, at the universities of Bologna, Barcelona, Leiden, Łódź, Dresden, Milano, Macerata, Sassari. Each of these audiences has provided fantastic feedback, helped me to refine my views, and to identify those bits of my research who ran the highest risk of being misunderstood. A special thanks goes to the organizers of the workshop “The Invention of Lying: Language, Logic and Cognition” and to the Lorentz Center in Leiden, for putting together such a great event and for their financial support.

If I survived my PhD and did so without losing my mind, friends and family are obviously to thank. I am grateful to all the people I lived with during these years: Kankana, Herbert, Oliver, Patrick, Isaac, Stephanie, Irene, Gerard, Lluc, and Carmen-MariCarmen. Similarly, I want to thank the ones who hosted me or that I hosted: Luca, Lucia, Luisa, Matteo, Francesco, Luigi, Ramon, Simone, Elena, il Cicogna and Timo. A special thanks goes to Giuseppina for the bacini, the melone and the coccole, for the frog nights, for washing the dishes and more importantly for keeping me alive; and to my whole family, by which I mean mamme, babbi, fratelli, zii, cugini e nonni, without whom this thesis would have never been written. Finally, I’d like to thank all those I forgot to mention, for allowing these acknowledgments to stay within the limit of these two pages.

# References

Adler, Jonathan E. 1997. Lying, Deceiving, or Falsely Implicating. *Journal of Philosophy.* 94: 435-452.

ALDRICH, VC. 1966. Telling, Acknowledging and Asserting. Analysis 27 (2): 53–56.

Alicke, Mark, David Rose & Dori Bloom. 2014. Causation, norm violation, and culpable control. In J. Knobe & S. Nichols (eds.). Experimental Philosophy (vol. 2), 229-250. New York: Oxford University Press.

Alston, WJ. 2000. Illocutionary Acts and Sentence Meaning. Ithaca: Cornell University Press

Anscombe, Elizabeth. 1981. Ethics, Religion and Politics. Oxford: Blackwell.

Aquinas. [ST] Summa Theologiae

Arico, Adam J., & Fallis, Don. 2013. Lies, Damned Lies, and Statistics: An Empirical Investigation of the Concept of Lying. Philosophical Psychology 26 (6): 790-816.

Augustine. [DDC] De Doctrina Christiana

Augustine. 1887 [DM]. De mendacio [On lying]. In Nicene and Post-Nicene Fathers, First Series, Vol. 3. ed. by Philip Schaff. Buffalo, NY: Christian Literature Publishing Co.

Austin, John L. 1961/2003. Performative Utterances. In J.L. Austin, Philosophical Papers. Oxford: Oxford University Press, 3rd edn.

Austin, John L. 1962/1975. *How to Do Things with Words,* Oxford: Oxford University Press, 2nd edn.

Bach, Kent and Robert M. Harnish. 1979. *Linguistic Communication and Speech Acts*. Cambridge, Mass.: MIT Press.

Bach, Kent. 1975. Performatives are statements too. *Philosophical Studies* 28: 229-236.

Bach, Kent. 2007. Knowledge in and out of context. In Campbell, J. K. & O’Rourke, M. (eds.), *Knowledge and skepticism*. Cambridge: MIT Press.

Bach, Kent. 2008. Applying Pragmatics to Epistemology. *Philosophical Issues* 18 (1): 68–88.

Barnes, John A. 1994. *A Pack of Lies: Towards A Sociology of Lying*. Cambridge, NY: Cambridge University Press.

Bazzanella, Carla, Caffi, Claudia & Sbisà, Marina. 1991. Scalar dimensions of illocutionary force. In Igor Ž. Žagar (ed.), *Speech acts: fiction or reality* (63-76). Antwerp/Ljubljana: IPrA Distribution Center for Yugoslavia and Institute for Social Sciences.

Bazzanella, Carla. 2009. Approssimazioni pragmatiche e testuali alla menzogna. (Pragmatic and textual approaches to the concept of lying). In *Tra Pragmatica e Linguistica testuale. Ricordando Maria-Elisabeth Conte*, Venier Federica (ed.), 67–90. Dell’Orso: Alessandria.

Beardsley, Monroe C. 1981. *Aesthetics: Problems in the Philosophy of Criticism*. Indianapolis: Hackett Publishing (Second Edition, 1958).

Belnap, Nuel D. 2000. Double Time References: Speech-act Reports as Modalities in an Indeterminist Setting. *Advances in modal logic* 3: 37-58.

Benton, Matthew A. 2013. Dubious Objections from Iterated Conjunctions. *Philosophical Studies* 162 (July 2011): 355–58. doi:10.1007/s11098-011-9769-3.

Benton, Matthew A. 2014. Gricean Quality. *Noûs* 50 (4): 689–703. doi:10.1111/nous.12065.

Benton, Matthew A. forthcoming. Lying, Belief, and Knowledge. In J. Meibauer, *The Oxford Handbook on Lying*. Oxford: Oxford University Press.

Boghossian , Paul A. 2003. The normativity of content. *Philosophical Issues* 13: 31–45.

Bok, Sissela. 1978. *Lying: Moral Choice in Public and Private Life*. New York NY: Random House.

Boyd, Kenneth. 2015. Assertion, Practical Reasoning, and Epistemic Separabilism. *Philosophical Studies* 172 (7): 1907–27. doi:10.1007/s11098-014-0378-9.

Brandom, Robert. 1983. Asserting. *Nous* 17 (4): 637–650.

Brandom, Robert. 1994. *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge: Harvard University Press.

Brown, Jessica. 2008. The Knowledge Norm for Assertion. *Philosophical Issues* 18(1): 89-103.

Caffi, Claudia. 1999. On Mitigation, *Journal of pragmatics* 31: 881-909.

Caffi, Claudia. 2007. *Mitigation*. London: Elsevier.

Cappelen, Herman. 2011. Against Assertion. In Brown, J. & Cappelen, H. (Eds.). *Assertion: New Philosophical Essays.* Oxford: Oxford University Press.

Carson, Thomas L. 1988. On the Definition of Lying: A reply to Jones and revisions. *Journal of Business Ethics* 7:509-514.

Carson, Thomas L. 2006. The definition of lying. *Nous* 2: 284–306.

Carson, Thomas L. 2009. Lying, deception, and related concepts. In *The philosophy of deception*, ed. Clancy Martin, 153–87. New York: Oxford University Press.

Carson, Thomas L. 2010. *Lying and Deception*. Oxford: Oxford University Press.

Carson, Thomas L., Richard E Wokutch and Kent F Murrmann. 1982. Bluffing in Labor Negotiations: Legal and Ethical Issues. *Journal of Business Ethics* 1(1): 13–22.

Carter, J Adam. 2017. Assertion, Uniqueness and Epistemic Hypocrisy. *Synthese* 194 (5): 1463–76. doi:10.1007/s11229-015-0766-5.

Carter, J. Adam. 2017. Assertion, Uniqueness and Epistemic Hypocrisy. *Synthese* 194 (5): 1463–76. doi:10.1007/s11229-015-0766-5.

Carter, JA. & EC. Gordon. 2011. Norms of Assertion: The Quantity and Quality of Epistemic Support. *Philosophia* 39(4): 615–635.

Chan, T., & Kahane, G. 2011. The trouble with being sincere. *Canadian Journal of Philosophy* 41(2): 215–234.

Chan, Timothy, and Guy Kahane. 2011. The Trouble with Being Sincere. *Canadian Journal of Philosophy* 41 (2): 1–13.

Chen, Rong, Chunmei Hu and Lin He. 2013. Lying between English and Chinese: An Intercultural Comparative Study. *Intercultural Pragmatics* 10.3: 375-401.

Chisholm, Roderick M. & Feehan, Thomas D. 1977. The intent to deceive. *Journal of Philosophy* 74(3): 143–159.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Coates, Jennifer. 1987. Epistemic modality and spoken discourse. *Transactions of the Philological Society* 85(1): 110–131.

Coffman, E. J. 2014. Lenient accounts of warranted assertability. In C. Littlejohn & J. Turri (Eds.), *Epistemic norms: new essays on action, belief and assertion* (33–59). Oxford: Oxford University press.

Cole, Shirley A. N. 1996. Semantic Prototypes and the Pragmatics of *Lie* Across Cultures. *The LACUS Forum* 23:475-83.

Coleman, Linda and Paul Kay. 1981. Prototype Semantics: The English Verb ‘lie’. *Language* 57: 26-44.

D’Agostini, Franca. 2012. *Menzogna* (Lying). Torino: Bollati Boringhieri.

Davidson, D. 2001. *Inquiries Into Truth and Interpretation.* New York: Oxford University Press

Davidson, Donald. 1985. Deception and Division. In J. Elster (ed.), *The Multiple Self*. Cambridge: Cambridge University Press.

Dawkins, Richard. 1989. *The Selfish Gene* (2nd ed)*.* Oxford: Oxford University Press.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of personality and social psychology*, *70*(5), 979.

DeRose, K. 2002. Assertion, Knowledge, and Context. *The Philosophical Review* 111(2): 167-203.

DeRose, Keith. 1991. Epistemic possibilities. *The Philosophical Review* 100(4): 581–605. DOI: 10.2307/2185175.

Donagan, Alan. 1977. *A Theory of Morality*. Chicago: Chicago University Press.

Douven, Igor. 2006. Assertion, Knowledge, and Rational Credibility. *Philosophical Review* 115 (4): 449–85. doi:10.1215/00318108-2006-010.

Douven, Igor. 2009. Assertion, Moore, and Bayes. *Philosophical Studies* 144 (3): 361–75. doi:10.1007/s11098-008-9214-4.

Dummett, M. 1981. Assertion, in *Frege: Philosophy of language*. London: Duckworth.

Dynel, Marta. 2011. A Web of Deceit: A Neo-Gricean View on Types of Verbal Deception. *International Review of Pragmatics* 3 (2): 137–137. doi:10.1163/187731011X610996.

Eco, Umberto. 1976. *Trattato di Semiotica Generale*. Milano: Bompiani. Translated as *A theory of Semiotics.* London: MacMillan.

[Egan, Andy & Weatherson, Brian (eds). 2011. *Epistemic Modality*. Oxford: OUP. DOI: 10.1093/acprof:oso/9780199591596.001.0001.](http://dx.doi.org/10.1093/acprof:oso/9780199591596.001.0001)

Egan, Andy, Hawthorne, John & Weatherson, Brian. 2005. Epistemic modals in context. In *Contextualism in Philosophy: Knowledge, Meaning and Truth*, Gerhard Preyer & Georg Peter (eds). Oxford: Clarendon Press.

Ekman, Paul. 1985. *Telling Lies: Clues to Deceit in the Marketplace, Marriage, and Politics.* New York: W.W. Norton.

Engel , Pascal. 2007. Belief and normativity? *Disputatio* , 2 ( 23 ), 153–77

Engel, Pascal. 2008. IN WHAT SENSE IS KNOWLEDGE THE NORM OF ASSERTION. *Grazer Philosophische Studien* 77 (1): 99–113.

Engel, Pascal. 2013. In defence of normativism about the aim of belief. In Chan, T. (ed.) *The Aim of Belief.* Oxford: Oxford University Press.

[Eriksson, John. 2011. Straight Talk: Conceptions of Sincerity in Speech. *Philosophical Studies* 153 (2): 213–234. doi:10.1007/s11098-009-9487-2.](http://doi:10.1007/s11098-009-9487-2)

Fadiman, Clifton.1985. *The Little, Brown Book of Anecdotes*. Boston: Little, Brown.

Falkenberg, Gabriel. 1988. Insincerity and disloyalty. *Argumentation* 2.1: 89-97.

Falkenberg, Gabriel. 1990. Searle on sincerity. In *Speech Acts, Meaning and Intentions.* Walter de Gruyter, Berlin, 129-145.

Fallis, Don. 2009. What Is Lying ? *The Journal of Philosophy* 106 (1): 29–56.

Fallis, Don. 2010. Lying and deception. *Philosophers’ Imprint*, (10) 1–22.

Fallis, Don. 2011. What liars can tell us about the knowledge norm of practical reasoning. *The Southern Journal of Philosophy* 49(4): 347–367. DOI: 10.1111/j.2041-6962.2011.00078.x.

Fallis, Don. 2012. Lying as a Violation of Grice’s First Maxim of Quality. *Dialectica* 66 (4): 563–581.

[Fallis, Don. 2013. Davidson was Almost Right about Lying. *Australasian Journal of Philosophy.* 91(2): 337–353. http://doi.org/10.1080/00048402.2012.688980](http://doi.org/10.1080/00048402.2012.688980)

[Fallis, Don. 2015. Are Bald-Faced Lies Deceptive after All? *Ratio* 28(1):81-96. http://doi.wiley.com/10.1111/rati.12055.](http://doi.wiley.com/10.1111/rati.12055)

Faulkner, P., 2007. What is Wrong with Lying? *Philosophy and Phenomenological Research*, 75: 524–547.

Faulkner, Paul. 2013. Lying and Deceit. In Hugh Lafollette (ed.), *International Encyclopedia of Ethics.* Hoboken, NJ: Wiley-Blackwell, 3101-3109.

Feehan, Thomas D. 1988. Augustine on Lying and Deception. *Augustinian Studies*, 131–39.

Feldman, R. S., Forrest, J. A., & Happ, B. R. 2002. Self-presentation and verbal deception: Do self-presenters lie more? *Basic and applied social psychology*, *24*(2), 163-170.

Foot, Philippa. 2001. *Natural Goodness*. Oxford: Oxford University Press.

Frankfurt, H. G. 1999. The Faintest Passion. In *Necessity, Volition and Love*, Cambridge: Cambridge University Press, 95-107.

Frankfurt, Harry G. 1986. On Bullshit. *Raritan Quarterly Review* 6:2, 81-100.

Frankish, Keith. 2009. Partial belief and flat-out belief. In Franz Huber & Christoph Schmidt-Petri (eds), *Degrees of Belief* (75–93). Berlin: Springer. DOI: 10.1007/978-1-4020-9198-8\_4.

Fraser, Bruce. 1980. Conversational mitigation. *Journal of Pragmatics* 4: 341-350.

Fraser, Bruce. 2010. Pragmatic Competence: The Case of Hedging. In *New Approaches to Hedging*, ed. G. Kaltenbo, W. Mihatsch & Schneider Stefan, 15–34. Emerald Group Publishing Limited.

Frege, G. 1918. Der Gedanke. *Beiträge Zur Philosophie Des Deutschen Idealismu*s. 58–77.

Frege, Gottlob. 1879. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens* [Concept Script, a formal language of pure thought modelled upon that of arithmetic]. Halle: L. Nebert. In *From Frege to Gödel: A Source Book in Mathematical Logic*, ed. by Jan van Heijenoort. Cambridge, MA: Harvard University Press.

Frege, Gottlob.1892. Über sinn und bedeutung, *Zeitschrift Für Philosophie Und Philosophische Kritik* [On sense and nominatum]100: 22–50.

Fried, Charles. 1978. *Right and Wrong*. Cambridge: Harvard

Fried, Charles. 1981. *Contract as Promise*. Cambridge, MA: Harvard University Press.

Friend, Stacie. 2008. Imagining Fact and Fiction. In *New Waves in Aesthetics*, edited by Kathleen Jones and Katherine Thomson-Jones. Palgrave Macmillan.

Friend, Stacie. 2014. Believing in Stories. In *Aesthetics and the Sciences of Mind*, edited by Greg Currie, Matthew Kieran, Aaron Meskin, and Jon Robson, 227–48. doi:10.1093/acprof:oso/9780199669639.003.0012.

Gale, Richard M. 1971. The Fictive Use of Language. *Philosophy* 46 (178) 324–40. doi:10.2307/3750012.

Garcìa-Carpintero, Manuel. 2004. Assertion and the Semantics of Force-Markers. In C. Bianchi (ed.) *The Semantics/Pragmatics Distinction*: 133–166*.* Stanford: The University of Chicago Press.

Garcìa-Carpintero, Manuel. 2013. Explicit Performatives Revisited. *Journal of Pragmatics* 49: 1–17.

Gerken, M. 2014. Same, same but different: the epistemic norms of assertion, action and practical reasoning. *Philosophical Studies* 168(3): 725-744.

Gerken, Mikkel. 2012. Discursive Justification and Skepticism. *Synthese* 189 (2): 373–94. doi:10.1007/s11229-012-0076-0.

Ginet, Carl, 1979. Performativity. *Linguistics & Philosophy* 3: 245-265.

Goldberg, S. 2015. *Assertion. On the philosophical significance of assertoric speech*. Oxford: Oxford University Press.

Gottwald, Siegfried. 2001. *A treatise on many-valued logics*. Baldock: Research Studies Press.

Graff, Gerald. 1980. *Poetic Statement and Critical Dogma*. Evanston: Northwestern University Press.

Graff,Gerald. 1979. *Literature against Itself. Literary Ideas in Modern Society.* Chicago: The University of Chicago Press.

Green, Mitchell S. 2000. Illocutionary Force and Semantic Content. *Linguistics and Philosophy*  23: 435–473.

Green, Mitchell S. 2007. *Self-Expression*. Oxford: Oxford University Press.

Green, Mitchell S. 2013. Assertions. In *Pragmatics of Speech Actions, Vol. II of the Handbook of Pragmatics*, II:0–33.

Greenough, Patrick. 2011. Truth‐Relativism, Norm‐Relativism, and Assertion. In *Assertion: New Philosophical Essays*, 197-232. doi:10.1093/acprof.

Grice, Herbert P. 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Griffiths, P. J. 2004. *Lying: An Augustinian theology of duplicity*. Wipf and Stock Publishers.

Grotius, Hugo. [RWP]. *The Rights of War and Peace*. F. W. Kelsey (trans.), Indianapolis: Bobbs-Merrill.

Gussenhoven, Carlos. 2002. Intonation and interpretation: phonetics and phonology. In *Proocedings of the Speech Prosody International Conference*, ed. by Stefan Sudhoff et al, 47-57. Aix en Provence: Laboratoire Parole et Langage.

Hájek, Petr. 1998. *Metamathematics of fuzzy logic*. Dodrecht: Kluwer.

Hardin, Karol J. forthcoming. Linguistic Approaches to Lying and Deception. In Meibauer, J. (ed.) *The Oxford Handbook on Lying*. Oxford: Oxford University Press.

Hare, Richard. 1952. *The language of morals*. Oxford: Oxford University Press.

Harris, Roy. 1978. The descriptive interpretation of performative utterances. *Journal of Linguistics* 14(2): 309-310.

Hartman, N. 1975. *Ethics*. Translated by S. Coit. Atlantic Highlands, NJ: Humanities Press.

Hawthorne, J. 2004. *Knowledge and Lotteries*. Oxford: Oxford University Press.

[Heal, J. 1974. Explicit Performative Utterances and Statements. *The Philosophical Quarterly* 24(95). http://doi.org/10.2307/2217715](http://doi.org/10.2307/2217715)

Hedenius, Ingemar. 1963. Performatives. *Theoria* 29: 115-36.

Hill, Christopher S, and Joshua Schechter. 2007. HAWTHORNE’S LOTTERY PUZZLE AND THE NATURE OF BELIEF. *Philosophical Issues* 17.

Hinchman, Edward S. 2013. Assertion, Sincerity, and Knowledge. *Noûs* 47 (4): 613–46. doi:10.1111/nous.12045.

Hindriks, F. 2007. The Status of the Knowledge Account of Assertion. *Linguistics and Philosophy* 30(3): 393– 406

Hindriks, F. 2009. Constitutive Rules, Language, and Ontology. *Erkenntnis* 71 (2): 253–275.

Hirst, Paul H. 1973. Literature and the Fine Arts as a Unique Form of Knowledge, *Cambridge Journal of Education* 3 (3): 118-132.

Holmes, Janet. 1984. Modifying illocutionary force. *Journal of Pragmatics* 8: 345–365. DOI: 10.1016/0378-2166(84)90028-6.

Holton, Richard. 2008. Partial Belief, Partial Intention. *Mind* 117: 27–58.

Hoye, Leo. 1997. *Adverbs and modality in English*. London, New York: Longman.

Huber, Franz & Christoph Schmidt-Petri (eds.). 2009. *Degrees of belief*. Springer.

Humberstone, Lloyd. 1992. Direction of Fit. *Mind* 101 (401).

Hume, David. [THN]*. A treatise on human nature.* NY: Oxford University Press.

Isenberg, Arnold. 1964. Deontology and the Ethics of Lying. *Philosophy and Phenomenological Research* 24: 463-480.

Jary, Mark. 2007. Are explicit performatives assertions? *Linguistics and Philosophy* 30(2): 207 - 234.

Johnson, Casey Rebecca. 2017. What Norm of Assertion? *Acta Analytica*, May. Springer Netherlands, 1–17. doi:10.1007/s12136-017-0326-3.

Juhl, P. D. 1980. *Interpretation. An Essay in the Philosophy of Literary Criticism*. Princeton: Princeton University Press

Kant, Immanuel. [*CPR*]. *Critique of pure reason*. Norman Kemp Smith. London: Macmillan

Kant, Immanuel. [GMM].Groundwork of the Metaphysics of Morals*.* Trans. by M. J. Gregor (trans.). In A. W. Wood and M. J. Gregor (eds.), *Immanuel Kant, Practical Philosophy.* Cambridge: Cambridge University Press.

Kant, Immanuel. [SRTL]. On a Supposed Right To Lie from Philanthropy. In Gregor, Mary J. (ed., trans.), *The Cambridge Edition of the Works of Immanuel Kant- Practical Philosophy*, 605-616. Cambridge: Cambridge University Press.

Keiser, Jessica. 2015. Bald-Faced Lies: How to Make a Move in a Language Game without Making a Move in a Conversation. *Philosophical Studies* 1 (2014). Springer Netherlands: 251–64. doi:10.1007/s11098-015-0502-5.

Kemp, Gary. 2007. Assertion as a Practice. In *Truth and Speech-Acts*, edited by D. Greimann and G. Siegwart, 106–29. Routledge.

Kenyon, Tim. 2003. CYNICAL ASSERTION: CONVENTION, PRAGMATICS, and SAYING ‘ UNCLE’. *American Philosophical Quarterly* 40 (3): 241–48.

Kenyon, Tim. 2010. ASSERTION AND CAPITULATION. *Pacific Philosophical Quarterly* 91: 352–68.

Kingsbury, Justine and Jonathan McKeown-Green. 2009. Definitions: Does Disjunction Mean Dysfunction? *Journal of Philosophy* 106:568-85

Klein, Peter. 1998. Certainty. In Edward Craig, *Routledge Encyclopedia of Philosophy* 264-267.

Knobe, Joshua. 2003. Intentional action and side effects in ordinary language. *Analysis* 279: 190–194.

Koethe, John. 2009. Knowledge and the Norms of Assertion. *Australasian Journal of Philosophy* 87 (4): 625–38. doi:10.1080/00048400802598660.

Kölbel, Max 2010. Literal Force: A Defence of Conventional Assertion. In Sarah Sawyer (ed*.), New Waves in Philosophy of Language*. London: Palgrave Macmillan, 108–37.

Kratzer, Angelika. 1981. The notional category of modality. In Hans-Jürgen Eikmeyer & Hannes Rieser (eds), *Worlds, Words, and Contexts*, 38–74. Berlin: De Gruyter.

Krishna, Daya. 1961. ‘Lying’ and the Compleat Robot. *The British Journal for the Philosophy of Science* 12 (46).

Kupfer, Joseph. 1982. The moral presumption against lying. *The Review of Metaphysics* 36(1): 103–126.

Kurthy, M., Lawford-Smith, H., & Sousa, P. (2017). Does ought imply can? *PloS one*, *12*(4), e0175206.

Kvanvig, J. 2009. Assertion, Knowledge, and Lotteries. In Greenough, P. & Pritchard, D. (eds.) *Williamson on Knowledge*. Oxford: Oxford University Press.

Kvanvig, Jonathan L. 2011. Norms of Assertion. In J. Brown and H. Cappelen (eds.) *Assertion: New Philosophical Essays.* doi:10.1093/acprof.

Labinaz, Paolo, and Marina Sbisà. 2014. Certainty and Uncertainty in Assertive Speech Acts. In *Communicating Certainty and Uncertainty in Medical, Supportive and Scientific Contexts*, Ilaria Riccioni, Carla Canestrari, Andrzej Zuczkowski, and Ramona Bongelli (eds.). Amsterdam: John Benjamins Publishing Company.

Labov, W. 1984. Intensity. In *Georgetown University Round Table on Language and Linguistics*, ed. by Deborah Schiffrin 43-70. Washington: Georgetown University Press.

Lackey, J. 2007. Norms of Assertion. *Nous* 41(4): 594–626.

Lackey, Jennifer. 2013. Lies and Deception: An Unhappy Divorce. *Analysis* 73 (2): 236–48. doi:10.1093/analys/ant006.

Lakoff, George. 1975. Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts. *Journal of Philosophical Logic* 2 (4): 458–508.

Lamarque, Peter & Stein Haugom Olsen.1994. *Truth, Fiction, and Literature. A Philosophical Perspective*. Oxford: Clarendon Press

Leland, Patrick R. 2015. Rational Responsibility and the Assertoric Character of Bald-Faced Lies. *Analysis* 75 (4): 550-554doi:10.1093/analys/anv080.

Lemmon, J.E. 1962. Sentences verifiable by their use. *Analysis* 12: 86-89.

Leonard, Henry S. 1959. Interrogatives, Imperatives, Truth, Falsity and Lies. *Philosophy of Science* 26: 172-186.

Levine, T. R. (ed.). 2014. *Encyclopedia of deception*. SAGE Publications.

Levinson, Stephen. C. 1983. *Pragmatics*. Cambridge: CUP.

Lewis, D. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8 (1):339--359.

Lewis, David. 1970. General Semantics. *Synthese* 22: 18-67

Lindley, T. Foster. 1971. Lying and Falsity. *Australasian Journal of Philosophy* 49: 152–157.

Lyons, John, 1977. *Semantics*. Cambridge: Cambridge University Press

MacCormick, N. (1983). WHAT IS WRONG WITH DECEIT? *Sydney Law Review.*

[MacFarlane, John. 2003. Epistemic Modalities and Relative Truth. Unpublished, Url= <http://johnmacfarlane.net/epistmod-2003.pdf*>.*](http://johnmacfarlane.net/epistmod-2003.pdf)

MacFarlane, John. 2003. Future Contingents and Relative Truth. *The Philosophical Quarterly* 53 (212): 321–36. doi:10.1111/1467-9213.00315.

MacFarlane, John. 2005. Making Sense of Relative Truth. *Proceedings of the Aristotelian Society*. 105 (1).

MacFarlane, John. 2011. What is Assertion? In J. Brown and H. Cappelen (eds.) *Assertion: New Philosophical Essays*, 79-96. Oxford: Oxford University Press.

MacIntyre, Alasdair. 1994. Truthfulness, Lies, and Moral Philosophers: What Can We Learn from Mill and Kant? *The Tanner Lectures on Human Values* 16: 307-61.

Mahon, James Edwin. 2007. A Definition of Deceiving. *International Journal of Applied Philosophy* 21 (2).

Mahon, James Edwin. 2008. Two Definitions of Lying. *International Journal of Applied Philosophy*, 211–230.

[Mahon, James Edwin. 2009. Why There are No Bald-Faced Lies. Paper presented at the Information Ethics Roundtable.https://itunes.apple.com/it/podcast/why-there-are-no-bald-faced-lies/id413143120?i=1000092195610&mt=2](https://itunes.apple.com/it/podcast/why-there-are-no-bald-faced-lies/id413143120?i=1000092195610&mt=2)

[Mahon, James Edwin. 2010. The definition of lying, and why it matters. Roundtable at the School of Info Resources and Library Sciences, https://itunesu.itunes.apple.com/feed/id413143120.](https://itunesu.itunes.apple.com/feed/id413143120)

[Mahon, James Edwin. 2011. Review of Lying and Deception by Thomas L. Carson, *Notre Dame Philosophical Reviews*. http://ndpr.nd.edu/news/24572-lying-and-deception-theory-and-practice/](http://ndpr.nd.edu/news/24572-lying-and-deception-theory-and-practice/)

Mahon, James Edwin. 2016. The Definition of Lying and Deception, in Zalta, E. N. (ed.), The Stanford Encyclopedia of Philosophy (Summer 2012 Edition).

Maitra, Ishani. 2011. Assertion, Norms, and Games. In Brown, J. & Cappelen, H. (eds.). *Assertion: New Philosophical Essays.* Oxford: Oxford University Press.

Mannison, Don S. 1969. Lying and lies. *Australasian Journal of Philosophy* 2(47): 132–144. DOI: 10.1080/00048406912341141.

Margolis, Joseph 1965, The Language of Art and Art Criticism. Detroit: Wayne State University Press.

Margolis, Joseph. 1962. “Lying Is Wrong” and “Lying Is Not Always Wrong”. *Philosophy and Phenomenological Research* 23: 414–418.

[Marsili, Neri. 2014. Lying as a Scalar Phenomenon: Insincerity along the Certainty-Uncertainty Continuum. In S. Cantarini, W. Abraham, and E. Leiss (eds.). *Certainty-Uncertainty – and the Attitudinal Space in between*, 153–173. Amsterdam: John Benjamins Publishing Company. 10.1075/slcs.165.09mar](http://dx.doi.org/10.1075/slcs.165.09mar)

Marsili, Neri. 2015. Normative accounts of assertion: from Peirce to Williamson, and back again. *Rivista Italiana di Filosofia del Linguaggio* (2).

Marsili, Neri. 2016. Lying by Promising. *International Review of Pragmatics*, 8(2), 271–313. http://doi.org/10.1163/18773109-00802005

Marsili, Neri. 2017. Lying and Certainty. In J.Meibauer, *The Oxford Handbook on Lying*. Oxford: Oxford University Press.

Marušić, Berislav. 2012. Belief and difficult action. *Philosopher's Imprint* *12*(18).

Marušić, Berislav. 2013. Promising against the Evidence. *Ethics* 123(2): 292–317.

McCammon, C. 2014. Representing Yourself as Knowing. *American Philosophical Quarterly* 51(2): 1–14.

McKenna, Robin. 2015. Assertion, Complexity, and Sincerity. *Australasian Journal of Philosophy* 93 (4) (October 2): 782–798.

Meibauer, Jörg . 2005. Lying and Falsely Implicating. *Journal of Pragmatics* 37(9): 1373–1399.

Meibauer, Jörg. 2011. On Lying: Intentionality, Implicature, and Imprecision. *Intercultural Pragmatics* 2 (8): 277–292.

Meibauer, Jörg. 2014. *Lying at the semantic-pragmatic interface*. Berlin: De Gruyter.

Meibauer, Jörg. 2014b. Bald-Faced Lies as Acts of Verbal Aggression. *Journal of Language Aggression and Conflict* 2 (1): 127–50. doi:10.1075/jlac.2.1.05mei.

Meibauer, Jörg. 2014c. A Truth That’s Told with Bad Intent: Lying and Implicit Content. *Belgian Journal of Linguistics* 28: 97–118. doi:10.1075/bjl.28.05mei.

Meibauer, Jörg. 2016. Understanding Bald-Faced Lies. *International Review of Pragmatics* 8 (2): 247–70. doi:10.1163/18773109-00802004.

Mele, Alfred. 1992. *Springs of Action.* Oxford: Oxford University Press.

Mikhail, J. 2011. *Elements of moral cognition: Rawls' linguistic analogy and the cognitive science of moral and legal judgment*. Cambridge University Press.

Mikkonen, Jukka. 2010. Implicit Assertions in Literary Fiction. *Proceedings of the European Society for Aesthetics* 2 (1960): 312–30.

Montaigne, Michel. 1595 [E]. *Essais.* Verdun: P. Villey et Saulnier.

Montminy, M. 2013. The Single Norm of Assertion. In A. Capone, F. Lo Piparo, & M. Carapezza (Eds.), *Perspectives on Pragmatics and Philosophy*: 35–52.

Moore, George. 1966. *Ethics* (2nd ed.), Oxford: Oxford University Press.

Moore, George. 1993. *Selected Essays*. London: Routledge.

Moran, Richard. 2005. Getting Told and Being Believed. *Philosophers’ Imprint.* 5(5).

Moran, Richard. 2005. Problems with sincerity. *Proceedings of the Aristotelian Society* 105: 325–345.

Nagel, T., 1970. *The Possibility of Altruism.* Princeton: Princeton University Press.

Newey, G., 1997. Political Lying: A Defense. *Public Affairs Quarterly* 11: 93–116.

Owens, David. 2008. Promising without Intending. *The Journal of Philosophy* 105(12): 737–755.

[Pagin, Peter. 2015. Assertion, In E. N. Zalta (ed.), *Stanford Encyclopaedia of Philosophy (Spring 2015 edition),* <http://plato.stanford.edu/archives/spr2015/entries/assertion/>.](http://plato.stanford.edu/archives/spr2015/entries/assertion/)

Pagin, Peter. 2016. Problems with Norms of Assertion. *Philosophy and Phenomenological Research* 93 (1): 178–207. doi:10.1111/phpr.12209.

Papafragou, A. 2006. Epistemic modality and truth conditions. *Lingua* 116(10), 1688–1702. http://doi.org/10.1016/j.lingua.2005.05.009.

Parsons, T. 1978. Review of John Woods: Logic of Fiction. *Synthese*, 39, 155-164.

Peirce, Charles S. [CP] *Collected Papers of Charles Sanders Peirce*, 8 vols. Edited by Charles Hartshorne, Paul Weiss, and Arthur W. Burks (Harvard University Press, Cambridge, Massachusetts, 1931–1958).

Peirce, Charles S. [MS] *The Charles S. Peirce Papers* (Cambridge: Harvard University Library, 1966, microfilm, 33 reels including supplement)

Pelling, Charlie. 2013. Assertion and Safety. *Synthese* 190 (17): 3777–96. doi:10.1007/s11229-012-0223-7.

Pelling, C. 2013b. Assertion and the Provision of Knowledge. *The Philosophical Quarterly* *63*(251). http://doi.org/10.1111/1467-9213.12013

Pepp, Jessica. forthcoming. *Truth Serum, Liar Serum, and Some Problems about Saying what You Think is False.* In E. Michaelson and A. Stokke (eds.), *Lying*. Oxford: Oxford University Press.

Peter Lombard. [SEN] *Libri Quattuor Sententiarum.*

Plantinga, Alvin. 1978. *The nature of necessity*. Oxford: Oxford University Press.

Plunze, Christian. 2001. Try to Make Your Contribution One that is True. *Acta Philosophica Fennica* 69:177-89.

Poggi, I., D’Errico, F., & Vincze, L. 2011. Discrediting moves in political debates. *In Proceedings of second international workshop on user models for motivational systems: the affective and the rational routes to persuasion*. Springer LNCS: 84-99.

Pollock, JL. 1982. *Language and Thought*. Princeton: Princeton University Press

Price, Huw. 1983. Does ‘Probably’ Modify Sense? *Australasian Journal of Philosophy* : 37–41.

Primoratz, Igor. 1984. Lying and the 'Methods of Ethics', *International Studies in Philosophy* XVI:35-57

Pritchard, D. 2013. Epistemic luck, safety, and assertion. In J. Turri & C. Littlejohn (Eds.), *Epistemic Norms: New Essays on Action, Belief, and Assertion*. Oxford: Oxford University Press.

Rawls, J. 1955. Two Concepts of Rules. *The Philosophical Review* 64(1): 3–32.

Rawls, John. 1981. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Récanati, François. 1987. *Meaning and Force: The Pragmatics of Performative Utterances.* Cambridge University Press, Cambridge

[Reed, Baron. 2008. Certainty. In *The Stanford Encyclopedia of Philosophy,* Edward N. Zalta (ed.), URL=http://plato.stanford.edu/archives/win2011/entries/certainty.](http://plato.stanford.edu/archives/win2011/entries/certainty)

Reichert, John. 1981. Do Poets Ever Mean What They Say? *New Literary History*. 13 (1):53-68.

Reichert, John.1977. *Making Sense of Literature*. Chicago: University of Chicago Press.

Reimer, Marga. 1995. Performative utterances: a reply to Bach and Harnish. *Linguistics and Philosophy* 18: 655-675.

Reimer, Marga. 2004. What malapropisms mean: A reply to Donald Davidson, *Erkenntnis* 60(3): 317– 334.

Rescorla, M. 2007. A Linguistic Reason for Truthfulness. In *Truth and Speech Acts*, edited by Dirk Greimann and Geo Siegwart, Routledge.

Rescorla, M. 2009. Assertion and Its Constitutive Norms. *Philosophy and Phenomenological Research* LXXIX (1): 98–130.

Reynolds, Steven L. 2002. Testimony, Knowledge, and Epistemic Goals. *Philosophical Studies* 110 (2): 139–61. doi:10.1023/A:1020254327114.

[Richard Ohmann. 1971. Speech Acts and the Definition of Literature. *Philosophy & Rhetoric* 4 (1). https://www.jstor.org/stable/pdf/40236740.pdf?refreqid=excelsior%3A59add992bef1d42defff9f8a31a4df7b.](https://www.jstor.org/stable/pdf/40236740.pdf?refreqid=excelsior%3A59add992bef1d42defff9f8a31a4df7b)

[Ridge, Michael. 2006. Sincerity and expressivism*. Philosophical Studies* 131: 487–510. http://doi.org/10.1007/s](http://doi.org/10.1007/s)

[Ridge, Michael. 2011. Reasons for Action: Agent-Neutral vs. Agent-Relative. In E. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition) URL = <https://plato.stanford.edu/archives/win2011/entries/reasons-agent/>.](https://plato.stanford.edu/archives/win2011/entries/reasons-agent/)

Ross, David. 1930. *The Right and the Good,* reprinted in 2002. Oxford: Oxford University Press.

Rowe, M. W. 1997. Lamarque and Olsen on Literature and Truth. *The Philosophical Quarterly* 47(188): 322-41.

Rutschmann, Ronja, and Alex Wiegmann. 2017. No Need for an Intention to Deceive? Challenging the Traditional Definition of Lying. PhilosoPhical Psychology. doi:10.1080/09515089.2016.1277382.

Ryan S. 2003. Doxastic compatibilism and the ethics of belief. *Philosophical Studies*. 114(1–2):47–79

Saul, Jennifer M. 2011. Just go ahead and lie. *Analysis* 72(1): 3–9.

Saul, Jennifer M. 2012. *Lying, Misleading, and what is Said: An Exploration in Philosophy of Language and in Ethics*. Oxford: Oxford University Press.

Sbisà, Marina. 2001. Illocutionary force and degrees of strength in language use. *Journal of Pragmatics* 33: 1791–1814. DOI: 10.1016/S0378-2166(00)00060-6

Sbisà, Marina. 2016. Varieties of Speech Act Norms. In *Dynamics and Varieties of Speech Actions*. doi:10.1017/CBO9781107415324.004.

Schaffer, J. 2008. Knowledge in the image of assertion. *Philosophical issues* 18(1).

Schopenhauer, Arthur. 1974. *Parerga e Paralipomena.* Translated by EFJ Payne (vol 2).

Schwyzer, H. 196). Rules and Practices. *Philosophical Review* *78*(4).

Searle, John R and Daniel Vanderveken. 1985. *Foundations of Illocutionary Logic*. Cambridge: Cambridge University Press.

Searle, John R. 1964. How to Derive Ought from Is. *The Philosophical Review* 73(1), 43-58

Searle, John R. 1965. What is a speech act? In Black, M.(ed.) *Philosophy in America*, 221-39. London, Allen & Unwin.

Searle, John R. 1969. *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139173438

Searle, John R. 1975. The Logical Status of Fictional Discourse. *New Literary History* 6 (2): 319–32.

Searle, John R. 1976. A Classification of Illocutionary Acts. *Language in Society* 5 (1): 1–23.

Searle, John R. 1989. How performatives work. *Linguistics & Philosophy* 12:535-558.

Searle, John R. 1995. *The Construction of Social Reality*. New York, The Free Press.

Searle, John R. 2007. Illocutionary acts and the concept of truth. In *Truth and Speech-Acts*, edited by D. Greimann and G. Siegwart, 106–29. Routledge.

Searle, John R., and Daniel Vanderveken. 2005. Speech acts and illocutionary logic. *Logic, Thought and Action*. Springer Netherlands. 109-132.

Seierstad, Asne. 2003. *A Hundred and One Days: A Baghdad Journal*, trans. by Ingrid Christophersen. New York: Basic Books.

Shah, Nishi .2003. How truth governs belief . *Philosophical Review* 112: 447–82 .

Shah, Nishi and J. David Velleman. 2005. Doxastic Deliberation. *Philosophical Review* 114: 497–534.

Sidgwick, Henry. 1981 [1874]. *The Methods of Ethics*. Indianapolis. IN: Hackett.

Sidney, S. P. 1595. *An Apology For Poetry (Or The Defence Of Poesy)*. Manchester: Manchester University Press (Revised and Expanded Second Edition, 2002).

Siegler, Frederick A. 1966. Lying. *American Philosophical Quarterly* 3: 128-136.

Simpson, David. 1992. Lying, Liars and Language, in *Philosophy and Phenomenological Research* 52: 623-639.

Smith, David. L. 2004. Why We Lie: The Evolutionary Roots of Deception and the Unconscious Mind. St. Martin's Press  
Unconscious Mind. St. Martin's Press.

Sorensen, Roy. 2007. Bald Faced Lies! Lying without the intent to deceive. *Pacific Philosophical Quarterly* 88: 251–264.

Sorensen, Roy. 2010. Knowledge-Lies. *Analysis* 70 (4) (August 7): 608–615. doi:10.1093/analys/anq072.

Sorensen, Roy. 2011. What lies behind misspeaking. *American Philosophical Quarterly* 48(4): 399-409.

Sosa, David. 2009. Dubious Assertions. *Philosophical Studies* 146 (August 2008): 269–72. doi:10.1007/s11098-008-9255-8.

Staffel, Julia. 2011. Reply to Sorensen, ‘knowledge-lies’. *Analysis* 71: 300–303.

Staffel, Julia. 2012. Can There Be Reasoning with Degrees of Belief? *Synthese* 2011: 1–20

Stalnaker, Robert. 1978. Assertion. In R. Stalnaker, *Context and content,* 78–95. Oxford: Oxford University Press.

Stalnaker, Robert. 1998. On the representation of context. In R. Stalnaker, *Context and content,* 96–114. Oxford: Oxford University Press.

Stanley, Jason. 2008. Knowledge and certainty. *Philosophical Issues* 18.1 35-5.

Stern R. 2004. Does ‘ought’ imply ‘can’? And did Kant think it does? *Utilitas*.16(1):42–61.

Stokke, Andreas. 2013. Lying and asserting. *Journal of Philosophy* 110(1): 33–60.

Stokke, Andreas. 2013b. Lying, Deceiving, and Misleading. *Philosophy Compas*s 8: 348–359.

Stokke, Andreas. 2014. Insincerity. *Noûs* 48 (3): 496–520. doi:10.1111/nous.12001.

Stokke, Andreas. 2016. Proposing, Pretending, and Propriety: A Response to Don Fallis. *Australasian Journal of Philosophy* (May): 1–6. doi:10.1080/00048402.2016.1185739.

Stone, Jim. 2007. Contextualism and Warranted Assertion. P*acific Philosophical Quarterly* 88: 92–113. doi:doi:10.1111/j.1468-0114.2007.00282.x.

Swanson, Eric. 2011. How not to theorize about the language of subjective uncertainty. In *Epistemic Modality,* Andy Egan & Brian Weatherson (eds), 249–269. Oxford: OUP. DOI: 10.1093/acprof:oso/9780199591596.003.0009.

Sweetser, Eve. 1987. The definition of lie: an examination of the folk models underlying a semantic prototype, in D. Holland and Q. Naomi (eds.), *Cultural Models in Language and Thought*, 3-66. Cambridge: Cambridge University Press.

Swift, Jonathan. 1710. The Art of Political Lying. *The Examiner,* 10*.*

Turri, Angelo and John Turri. 2015. The Truth about Lying. *Cognition.* 138: 161–168.

Turri, J. 2013. The Test of Truth: An Experimental Investigation of the Norm of Assertion. *Cognition* 129 (2): 279–291.

Turri, John, and Peter Blouw. 2015. “Excuse Validation: A Study in Rule-Breaking.” *Philosophical Studies* 172 (3): 615–34.

Turri, John. 2010. Epistemic Invariantism and Speech Act Contextualism. *Philosophical Review* 119 (1): 77–95.

Turri, John. 2011. The Express Knowledge Account of Assertion. *Australasian Journal of Philosophy* 89 (1): 37–45.

Turri, John. 2013. Knowledge and Suberogatory Assertion. *Philosophical Studies* 167 (3): 557–67.

Urmson, J. O. 1976. *Fiction* American Philosophical Quarterly 13 (2): 153-157.

Van der Henst, Jean-Baptiste, Laure Carles & Dan Sperber. 2002. Truthfulness and Relevance in Telling The Time. *Mind and Language*, 17 (5) (November): 457–466.

Van der Schaar, Maria. 2011. Assertion and Grounding: A Theory of Assertion for Constructive Type Theory. *Synthese* 183: 187–210. doi:10.1007/s11229-010-9758-7.

Van Inwagen, Peter. 1977. Creatures of Fiction. *American Philosophical Quarterly* 14 (4): 299–308.

Vanderveken, D. 1990. *Meaning and Speech Acts: Volume 1, Principles of Language Use* (Vol. 1). Cambridge University Press.

Vanderveken, Daniel. 1980. Illocutionary logic and self-defeating speech acts. In J.R. Searle, F. Kiefer and M. Bierwisch, *Speech act theory and pragmatics,* 247-272. Amsterdam: Springer.

[Viebahn, E. 2017. *Non-literal Lies*. Erkenntnis. http://doi.org/10.1007/s10670-017-9880-8](http://doi.org/10.1007/s10670-017-9880-8)

Vlach, F. 1981. Speaker's meaning. *Linguistics and Philosophy.* 4 (3),359-391.

Von Fintel, Kai, & Anthony S. Gillies. 2008. CIA leaks. *Philosophical review*, 117(1), 77-98. doi: 10.1215/00318108-2007-025

Vrij, A. 2008. *Detecting Lies and Deceit. Pitfalls and Opportunities*. 2nd ed, 1st ed. 2000. Chichester: Wiley

Walton, Kendall. 1990. *Mimesis as make-believe: On the foundations of the representational arts*. Harvard University Press.

Watson, Gary. 2004. Asserting and Promising. *Philosophical Studies* 117(1): 57–77.

Wedgwood, Ralph. 2002. The Aim of Belief. *Philosophical Perspectives*, 16, 267–97.

Weiner, Matthew. 2005. Must We Know What We Say? *Philosophical Review* 114 (2): 227–251.

Whiting, Daniel. 2010. Should I Believe the Truth? Dialectica 64 (2): 213–24. doi:10.1111/j.1746-8361.2009.01204.x.

Whiting, Daniel. 2012. Stick to the Facts: On the Norms of Assertion. *Erkenntnis* 78 (4): 847–867.

Whiting, Daniel. 2013. The Good and the True (or the Bad and the False). *Philosophy* 88: 219-242

Wiegmann, Alex, Jana Samland & Michael R. Waldmann, 2016. Lying despite telling the truth. *Cognition* 150: 37-42.

Williams, Bernard. 1966. Consistency and realism, *Proceedings of the Aristotelian Society*, 40: 1-22.

Williams, Bernard. 1985. *Ethics and the Limits of Philosophy*. Cambridge, MA: Harvard University Press.

Williams, Bernard. 2002. *Truth and Truthfulness: An Essay in Genealogy*. Princeton NJ: Princeton University Press .

Williams, John N. 1996. Moorean Absurdities and the Nature of Assertion. *Australasian Journal of Philosophy* 74 (September 2014): 135–49. doi:10.1080/00048409612347111.

Williamson, Timothy. 1996. Knowing and Asserting. *The Philosophical Review* 105, 4, 489–523.

Williamson, Timothy. 2000. Assertion. In *Knowledge and Its Limits*, 238-270. Oxford University Press.

Wilson, Deirdre & Dan Sperber. 2002. Truthfulness and Relevance. *Mind* 25: 1–41.

Wittgenstein, Ludwig. [PI]. *Philosophical investigations.* 4th ed., trans. by Anscombe, Hacker & Schulte. Wiley-Blackwell

Wood, D., 1973. Honesty. In A. Montefiore (ed.), *Philosophy and Personal Relations: An Anglo-French Study*, London: Routledge, 192–218.

Wright, C. 1992. *Truth and Objectivity*. Cambridge, Mass.: Harvard University Press.

Xu, Fen, Yang C. Luo, Genyue Fu, and Kang Lee. 2009. Children’s and adults’ conceptualization and evaluation of lying and truth-telling. *Infant and Child Development* 18(4): 307–322.

Yalcin, Seth. 2007. Epistemic Modals. *Mind*, *116*(464), 983–1026. http://doi.org/10.1093/mind/fzm983

Yalcin, Seth. 2011. Nonfactualism about epistemic modality. In Andy Egan & B. Weatherson (eds.), *Epistemic Modality*. Oxford University Press.

1. Luke Lythgoe and Hugo Dixon, “EU-bashing stories are misleading voters – here are eight of the most toxic tales”, *The Guardian* (Thursday 19 May 2016). [↑](#footnote-ref-1)
2. Ashley Kirk, “EU referendum: The claims that won it for Brexit, fact checked”, *The Telegraph* (13-03-2017); Jon Stone, *The Independent,* “Nigel Farage backtracks on Leave campaign's '£350m for the NHS' pledge hours after result” (24 June 2016). [↑](#footnote-ref-2)
3. Maria Konnikova, “Trump’s Lies vs. Your Brain”, *Politico Magazine* (January 2017); “Donald Trump’s File at *Politifact* (url: <http://www.politifact.com/personalities/donald-trump/>); [↑](#footnote-ref-3)
4. Dana Liebelson, Jennifer Bendery, Sam Stein, “Donald Trump Made Up Stuff 71 Times in an Hour”, *Huffington Post* (30-6-2017). [↑](#footnote-ref-4)
5. David Leonhardt & Stuart A. Thompson, “Trump’s Lies”, *New York Times* (23-6-2017). [↑](#footnote-ref-5)
6. Leonard Cassuto, “Big trouble in the world of 'Big Physics'”, *The Guardian* (18-9-2002). [↑](#footnote-ref-6)
7. An exception might apply to the initial propagator(s) of the theory, who could have invented it with deceptive intentions. For the purpose of the present discussion, let us call a *conspiracy theorist* onlysomeone who *genuinely believes* in a false and implausible conspiracy theory. [↑](#footnote-ref-7)
8. Craig Silverman, “This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook”, *Buzzfeed* (16-11-2016); Jason Tanz, “Journalism fights for survival in the post-truth era”, *Wired*, (14-2-2017). [↑](#footnote-ref-8)
9. Anti-vaccine body count, at http://www.jennymccarthybodycount.com/ [↑](#footnote-ref-9)
10. Contrary to what is argued by some authors (Siegler 1969: 129n), Aquinas did not require a deceptive intention: “The desire to deceive belongs to the perfection of lying, but not to its species, as neither does any effect belong to the species of its cause” (Aquinas ST, II, IIae, q110, a2). Augustine’s case is more complex: the traditional interpretation is that he endorsed the intention to deceive condition (e.g. Siegler 1969: 129n; Feehan 1988; 135-8) but according to Griffiths (2004: 30) this view is misguided, and there are better reasons to think that he was partisan of a “non-deceptionist” definition of lying. I believe that an accurate and charitable reading of Augustine’s work can only support an interpretation lying between these two extremes: Augustine simply did not settle the question as to whether a deceptive intention is required for lying or not. [↑](#footnote-ref-10)
11. These are generally biologists (e.g. Smith 2000, Dawkins 1989:64) and psychologists (e.g. Ekman 1985:26-8, Vrij 2008), who often treat the verb “lying” as equivalent to “intentionally deceiving”. [↑](#footnote-ref-11)
12. Some might argue that some animal signalling can count as lying. It is unclear whether non-human animals can meet the statement condition, which requires the ability to communicate in a language that assigns meanings to expressions. But if some animals can communicate in a language, it could be argued that they can also lie. Some cases of animals that purportedly used language to lie are reported in the literature. For instance, Fouts & Mills (1997:156) report the following dialogue with Lucy, a chimpanzee trained to speak in American Sign Language:

    *Fouts: What that? [indicating a pile of chimpanzee feces on the floor]  
    Lucy: What that?  
    Fouts: You known. What that?  
    Lucy: Dirty dirty.  
    Fouts: Whose dirty dirty?  
    Lucy: Sue. [a reference to Sue Savage-Rumbaugh, a graduate student of Fouts]  
    Fouts: It not Sue. Whose that?  
    Lucy: Roger!  
    Fouts: No! Not mine. Whose?  
    Lucy: Lucy dirty dirty. Sorry Lucy.* [↑](#footnote-ref-12)
13. Augustine (DM, 3.3), Aquinas (ST, II-II, q.110, a1), Kant (1797), Leonard (1959:182), Isenberg (1964:466; 1974), Lindley (1971), Mannison (1969:138), Chisholm & Feehan (1977), Kupfer (1982:104), Adler (1997), Williams (2002), Mahon (2008), Fallis (2009). Other authors endorse weaker positions. Carson (1982:16; 2006:284; 2010:39) and Saul (2012) provide a definition that does not require falsity, but both suggest that the definition could be strengthened to include such a requirement. Siegler (1966:132) suggests that falsity is necessary for telling a lie but not for lying; less controversially, Coleman & Kay (1981:28) argue that falsity is necessary for *prototypical* lying. [↑](#footnote-ref-13)
14. This label can itself be deceptive, as it may be interpreted as suggesting that these accounts require successful deception. Only intended deception is required: “deceptionist” should be taken to be a shorthand for “based on the intent to deceive condition”. [↑](#footnote-ref-14)
15. Cf. Sidney (1595), Austin (1962/1975), Margolis (1965:158-9), Gale (1971:324-339), Ohmann (1971:11-14), Searle (1975), Plantinga (1978:161-2), Van Invagen (1977:301), Parsons (1978:158), Beardsley (1981:419-23), Lamarque & Olsen (1994: 321), Mikkonen (2010). For an opposite view, cf. Hirst (1973), Urmson (1976), Juhl (1980), Graff (1979, 1980), Reichert (1977, 1981), Rowe (1997); cf. also Currie (1990:48-9), Walton (1990:79), Friend (2008; 2014). [↑](#footnote-ref-15)
16. Perhaps a way to escape this criticism is to contend, à la Lewis (1978), that fictional utterances are only claimed to be true in the fictional world, and hence they do not count as lies because they do not satisfy the insincerity condition for lying. According to this analysis, George’s statement “Scots begun to wear the kilt at the beginning of the XIst century” means something different, namely *“In the world of this fiction*, Scots begun to wear the kilt at the beginning of the XIst century”. Since George does not believe the fictionally-quantified proposition to be false, his statement, though deceptive, is not a lie. At least three difficulties can be mentioned for this approach. First, this strategy is only viable if the Lewisian analysis of fictional utterances is correct, but there are known difficulties with it (e.g. Currie 1990:62-81). Second, this analysis has been defended as an account of what it means to be true-in-fiction, and it is not clear why we would accept it as an account of what is *literally said* by a fictional utterance – so that the deceptionist would have to tell us something about how the fictional operator “in the world of fiction, …” comes to be part of the *literal* content of the statement. Third, and relatedly, if we interpret fictional statements along these lines we have undesirable consequences; for instance, that *unreliable fictions* contain *genuine lies* rather than merely deceptive statements. Unreliable fictions are fictions in which the narrator sometimes states propositions that later turn out to be false in the story (cf. Booths 1961: 158–9). When this happens for a proposition *p*, all the deceptionist conditions are met: the author is stating that *p* is true *in the world of fiction*, believes that *p* is false in that world, and his utterance is intended to deceive. The account can perhaps be refined to deal with these objections. But absent any refinements, this view fails to make the right predictions; and even if these refinements can be produced, this strategy will not be able to deal correctly with other non-assertoric falsehoods, like the Ironic Double Bluff case that I am about to discuss. [↑](#footnote-ref-16)
17. Since Jack already suspects Neil to be an undercover cop, one could argue that Neil is only attempting to alter Jack’s degree of confidence in the truth of the proposition, and that this is not enough to satisfy IDC. However, many philosophers agree that in these cases it is appropriate to talk about intended deception (Krishna 1961: 147; Chisholm & Feehan 1977: 145; Fallis 2009:45, 2011; Staffel 2011; 2012 – but see Carson 2009: 181), and in Ironic Double Bluff it seems unproblematic to claim that Neil is trying to deceive Jack. [↑](#footnote-ref-17)
18. One could argue that jocular lies are jokes rather than lies, and that consequently the term “jocular lies” is just a broad way of speaking. Alternatively, one could contend that jocular lies do aim at deceiving, even if deception is meant to be revealed later. I tend to be sympathetic to this view: I presented the case of jocular lies for its historical relevance, not as a genuine counterexample to deceptionist accounts. However, it is worth reminding that it was regarded as a genuine counterexample: “Yet not every lie is a cause of deception, since no one is deceived by a jocose lie; seeing that lies of this kind are told, not with the intention of being believed, but merely for the sake of giving pleasure.” (Aquinas ST, II, IIae, q110, a3) [↑](#footnote-ref-18)
19. Carson (1988: 510; 2006: 291) offers a thorough discussion of the difference between side effects and intended effects, and of how this relates with the debate about the IDC (cf. also Chisholm & Feehan 1977: 148n6). [↑](#footnote-ref-19)
20. A previous discussion of bald-faced lies was provided by Kenyon (2003: 242), that terms them “cynical assertions”. He defines them as follows: “[A] Cynical speaker S produces an utterance under conditions such that: (i) S is insincere; (ii) S’s audience correctly believes (i); and both S and S’s audience have higher-order attitudes of these sorts towards the other’s attitudes at every reasonable higher-order level. The slightly convoluted phrasing of this last clause indicates that, were it extended, the list of conditions would go: (iii) S knows (ii); (iv) S’s audience believes (iii); (v) S knows (iv)...; and so forth, to any level that either S or S’s audience might bother to formulate such an attitude”. [↑](#footnote-ref-20)
21. By saying that *p* is common knowledge within a group of agents I mean that each member of the group knows that *p*, and knows that each member knows that *p*, and knows that each member knows that each member knows that *p*, and so forth *ad infinitum*. [↑](#footnote-ref-21)
22. Sorensen (2007, 2010) and Arico & Fallis (2013) alone offer several examples. Meibauer (2016) offers a nice overview of the most successful examples in the literature. [↑](#footnote-ref-22)
23. Many deceptionist accounts, however, do not require that the proposition be asserted (Isenberg 1973:248, Primoratz 1984:54, Newey 1997, Williams 2002, Mahon 2008, Lackey 2013). Replies of this kind are thus only able to defend deceptionist accounts that also require that the proposition be asserted (as Chisholm & Feehan 1977, Fried 1978:57, Simpson 1992, Williams 2002, Faulkner 2007, Meibauer 2005; 2014). [↑](#footnote-ref-23)
24. The most prominent example of a bald-faced lie is certainly Cheating Student (e.g. Sorensen 2007:254, Fallis 2009; 2010; 2014:2, Dynel 2011:150, Lackey 2013:3, Arico & Fallis 2013:792, Meibauer 2014; 2014b:103; 2014c:137). The most prominent example of a coercion-lie is the variant of Coerced Witness in which no CCTV footage is shown (originally in Siegler 1996:129, often quoted from Carson 2006:289, it is found in Jones 1986:348, Carson 1988:509, Sorensen 2007:254, Mahon 2008:225, Dynel 2011:150, Lackey 2013:3, Fallis 2013:343, Keiser 2015, Meibauer 2011, 2014c:132, 2016b). The difference is relevant, as this variant is obviously an assertion also on Leland’s account, given that the audience has no way to suspect of the coercion. Compare the prominence of the mentioned examples Leland’s targets, that are barely mentioned in the literature. The case of the scared witness example contradicted by CCTV footage has only been quoted in other three works (Fallis 2009, Saul 2012:9, Stokke 2013). The totalitarian state example is only quoted once in the literature, and in this unique case it is acknowledged to be controversial (Saul 2012:9). [↑](#footnote-ref-24)
25. I am simplifying Keiser’s argument here, but such simplifications can be ignored for the purpose of the present discussion. [↑](#footnote-ref-25)
26. The term R-intention (reflexive intention) is a term of art frequently used by neo-Griceans scholars. Broadly, for A to R-intend that B performs an action *f* is for A to intend that B *f*s at least partly in virtue of the recognition of A's intention that B *f*s (cf. Bach and Harnish 1979). [↑](#footnote-ref-26)
27. For criticism of the definition of assertion employed by Keiser, see Aldrich (1966), Searle (1969), Vlach (1981), Davis (1999), Alston (2000:44-50), Green (2007:75-82), McFarlane (2011). [↑](#footnote-ref-27)
28. Carson (2006, 2010) and Saul (2012) suggest that a further condition might be required, namely that the asserted proposition be actually false – but neither commits to this further requirement (cf. IV.3.8.3). [↑](#footnote-ref-28)
29. Admittedly, Carson does not adopt the view that lying requires asserting, but suggests (2006:300-1) that he is sympathetic to the idea. I leave out of the discussion Sorensen’s (2007) account because he does not provide a detailed account of assertion, and Saul’s (2012) because it is based on Carson’s. [↑](#footnote-ref-29)
30. Here Fallis is referring to Grice’s analysis of meaning (1989) – to be more specific, to Grice’s analysis of the expression “In uttering U, S means that *p*”. The condition that in order to mean that *p* the speaker must intend to make the audience believe that *p* is found in Grice (1989: 219). However, this analysis is subject to known counterexamples, and has been further refined by Grice, cf. (1989: §5,6,14,18). [↑](#footnote-ref-30)
31. This figure is based on a real survey, conducted in 2011: <http://www.publicpolicypolling.com/pdf/2011/PPP_Release_National_ConspiracyTheories_040213.pdf> [↑](#footnote-ref-31)
32. See Peirce (CP 2.335, 2.535, 5.542, 8.115; MS 70.52); Frege (1918:22), Moore (1912:125), Black (1952: 31), Hare (1952:13,19–20,168–99), Searle (1969:60, 64–8), Bach and Harnish (1979), Unger (1975: 257), Wright (1992:14), Williams (1996:136), Green (2007:70-83), Rescorla (2009: 108–14). Since representing yourself as believing *p* does not require that you actually believe that *p,* this parallelism works only if we also understand the notion of expressing a belief as not requiring that you actually believe that *p ­*– as most (but not all) philosophers do. [↑](#footnote-ref-32)
33. Stalnaker (1978) is aware of this, and specifies that his proposal is not a *definition* of assertion, but rather a necessary but not sufficient condition for asserting that *p*. [↑](#footnote-ref-33)
34. As a reminder, by saying that *p* is common knowledge within a group of agents I mean that each member of the group knows that *p*, and knows that each member knows that *p*, and knows that each member knows that each member knows that *p*, and so forth *ad infinitum* [↑](#footnote-ref-34)
35. To be entirely precise, even if virtually every author in the literature agrees that the insincerity condition must be satisfied for a statement to count as a lie, a minority of authors (including myself) have proposed amendments to this condition. For a review of these proposals, cf. IV.2-5. [↑](#footnote-ref-35)
36. E.g. Hedenius (1963), Lewis (1970), Bach (1975), Ginet (1979), Bach & Harnish (1979:§10; 1992), Green (2005). [↑](#footnote-ref-36)
37. E.g. Austin (1962/1975), Harris (1977), Searle & Vanderveken (1985), Searle (1989), Reimer (1995), Jary (2000). [↑](#footnote-ref-37)
38. The square brackets indicate the truth-evaluable content of the sentence, where that content does not appear in a truth-evaluable form in the performative utterance. My proposed identification of the truth-evaluable content of performative utterances is not arbitrary, but rather follows non-descriptivist standards (cf. Searle 1989). [↑](#footnote-ref-38)
39. (1) I assure you that (1\*) I stole the cake

    (2) I promise that (2\*) I will clean the toilet today

    (4) I order you to clean the toilet (4\*) [that you will clean the toilet]

    (5) Assume that (5\*) the defendant is guilty…

    (6) I bet that (6\*) Miklos committed the murder! [↑](#footnote-ref-39)
40. (4) I order you to clean the toilet (4\*) [that you will clean the toilet]

    (5) Assume that (5\*) the defendant is guilty…

    (6) I bet that (6\*) Miklos committed the murder! [↑](#footnote-ref-40)
41. Perhaps in uttering (6a) the speaker is not representing himself as having a partial belief in the proposition (6\*), but rather an outright belief in the proposition (6a). I have at least three qualms with this reply. First, it seems that this explanation misses something important about (6a): there is clearly a sense in which (6a) expresses a partial belief; given that ACF3 is an account of assertion in terms of belief-expression, it seems that it cannot ignore this linguistic datum altogether. Second, if we adopt a non-descriptivist account when we apply ACF3 to performative utterances (according to which “I promise” is not part of the content of the utterance “I promise that *p*”, but rather indicates its force), it seems *ad-hoc* to then adopt a descriptivist account of epistemic modals (according to which “probably” is part of the content of “probably *p”,* even if it affects its force) – so that something should be said to justify this asymmetry. Third, there are independent reasons not to adopt a descriptivist account of epistemic modals when it comes to modelling lies, and choosing this strategy would lead to further problems, that I discuss in chapter IV.4.5.3. [↑](#footnote-ref-41)
42. There is agreement that the content must be believed to be false, but not that it must be identified in a non-descriptivist fashion. [↑](#footnote-ref-42)
43. According to a standard view, the direction of fit of a speech act (or propositional attitude) with content *p* indicates the way in which *p* has to conform (match, fit) to the world of the utterance (or of the propositional attitude). There are two main directions of fit: *word-to-world* and *world-to-word*. Assertions and beliefs have *word-to-world* direction of fit: the ‘word’ (the content of the assertion or the belief) purports to fit the ‘world’ (an independently existing state of affairs). Commands and desires, by contrast, have *world-to-word* direction of fit: they are satisfied when the ‘world’ is rearranged so that it fits the ‘word’. [↑](#footnote-ref-43)
44. Perhaps these philosophers are wrong, and there can be such a thing as unintentional lying. If so, it does not follow that my definition is wrong: it just follows that there are some lies that fall out of my envisaged *explanandum.* For a discussion of a putative case of *unintentional* lying, see the case of the confused politician in Carson (2006:296) and Arico and Fallis (2013). [↑](#footnote-ref-44)
45. Alston (2000: 120) offers an even more precise formulation: “S explicitly presents the proposition that p, or S is uttered as elliptical for a sentence that explicitly presents the proposition that *p*”. This condition is meant to deal with assertions whose content is borrowed from another sentence, as it happens in replies to questions. For instance, in the conversational interaction “A: Did you kill your neighbour’s cat? B: Yes”, B’s reply is elliptical for “Yes, I killed the cat”; roughly put, the reply “borrows” its propositional content from the question. [↑](#footnote-ref-45)
46. This formulation of Alston’s notion of responsibility (in his terminology, “R’ing”) is just one of the many presented in his (2000), and it is not the one that Alston eventually takes on board. Alston’s preferred notion of R’ing implies that it is permissible for the speaker to assert *p* only if *p* is true. Since I am not willing to build this requirement into the definition of assertion (cf. chapter VI), I preferred to refer to his formulation of the notion of R’ing that does not carry this implication. For more on Alston’s notion of R’ing, cf. Alston (2000:54-64). [↑](#footnote-ref-46)
47. Following standard notation conventions, here ‘F’ indicates the illocutionary force of the speech act, and ‘p’ indicates its propositional content. [↑](#footnote-ref-47)
48. One might worry that the Deep Throat example poses a threat also to my view, despite its being designed to attack ACS. It could be argued that, in announcing “If you repeat it, I will deny it”, Deep Throat is making it explicit not only that he does not want the proposition to become common ground, but also that he is refusing to commit himself to the truth of the proposition. This worry, however, is unfounded. In the Deep Throat example, the relevant responsibilities arise *in virtue of* Deep Throat’s relevant utterance– he can appropriately be reproached for having said something false, and appropriately challenged for what he said*.* Furthermore, according to the account of commitment I delineated in section 3.1, it is always possible for the speaker to discharge his commitment by retracting the assertion. In announcing that he would be willing to make such a move were his assertion to be challenged, Deep Throat is not refusing to accept discursive responsibility: he is merely disclosing that, if challenged, he will make a move that is within his conversational rights, namely retract the assertion – something that is surely uncommon, but not clearly in contradiction with one’s discursive responsibilities. [↑](#footnote-ref-48)
49. Similar problems apply to assertions mitigated by epistemic modals. For a discussion of how epistemic modals can affect whether or not a speaker has asserted a proposition, cf. IV.4.5.3. [↑](#footnote-ref-49)
50. According to a standard notation, I use ↓ to indicate the word-to-world direction of fit (the point of the utterance is to get the propositional content to match the world of the utterance), ↑for world-to-word (the point of the utterance is to get the world of the utterance to match the propositional content), ↕ for both, and ∅ for null direction of fit. S indicates the speaker, H the hearer. [↑](#footnote-ref-50)
51. One might wonder whether *denials* represent a counterexample to my account. We have seen that in these cases the speaker is committed to the *opposite* of what is said. The worry is that, given the mismatch between the proposition expressed and the proposition to which the speaker is committing, my definition is unable to capture denials as asserted (and therefore as lies). But this is not a problem unless we have a reason to interpret the notion of ‘expressing’ in a narrow way (as prohibiting such mismatch). Searle & Vanderveken do not seem to interpret the notion in this way. They seem to think that the performative verb “deny” is an *illocutionary force indicating device* (IFID)that indicates that the speaker is performing an assertion with content “¬*p”* (for a brief, formal discussion of this point, cf. Searle & Vanderveken 1985:183). Here an IFID is “any element of a natural language which can be literally used to indicate that an utterance of a sentence containing that element has a certain illocutionary force or range of illocutionary forces” (1985:2). Lastly, if we do not allow for a ‘loose’ interpretation of the notion of content of performative utterance, also every other account will not be able to count denials– so *denials* would also be a counterexample to every other definition. [↑](#footnote-ref-51)
52. As a matter of fact, some authors are convinced that declarations are assertions. Searle and Vanderveken (1985:177) claim that “the illocutionary force of declaration illocutionary entails that of assertion”; in other words, they believe that declarations are always assertions. For reasons that I am about to discuss, even if my account relies on Searle and Vanderveken’s framework (from which I rarely depart), I do not endorse this thesis unconditionally, as I find it controversial. [↑](#footnote-ref-52)
53. The longer sections in this chapter are based on material that I have already published elsewhere. Section 2 draws on Marsili (2016); section 3 is an almost literal excerpt from the same paper. Section 4 is based on material from two different papers, Marsili (2014) and Marsili (2017). [↑](#footnote-ref-53)
54. For a review of different conceptions of insincerity, cf. Eriksson (2011). [↑](#footnote-ref-54)
55. In a recent paper, Jessica Pepp (forthcoming) mentions some further difficulties that may arise in this respect (cf. also Chan & Kahane 2011, Stokke 2014). Considering that Pepp’s problems seem to emerge from problems affecting theories of reference in general rather than insincerity in particular, and given that this chapter is concerned with problems that are distinctive of defining lying, I will leave the discussion of these subtle counterexamples for another time. [↑](#footnote-ref-55)
56. It is debatable, for instance, whether a question or request always expresses a desire. For the purpose of the dissertation, however, we can leave this question aside. Independently of which attitude a question expresses, what matters is that we can plug the correct characterisation of question into the general model that I am adopting. [↑](#footnote-ref-56)
57. This view has been defended, under different guises, by Hare (1952:13,19–20,168–99), Searle (1969:60, 64–8), Wright (1992:14), Williams (1996:136), Moran (2005b), Green (2007:70-83). [↑](#footnote-ref-57)
58. One might be wary of the chosen scope of the negation in INS, as we could that S is Ψ¬(p) rather than requiring that S is not in Ψ(p) In section 5, I will address this kind of worry, and present an alternative version of this condition. [↑](#footnote-ref-58)
59. This view is extremely influential in the philosophical literature on the nature of social obligations. Many authors (sometimes referred to as ‘information-interest’ theorist) take promising’s main function to be informing the promisee of what the promisor is going to do (Sidgwick 1981:442–44, Anscombe 1981:18, Rawls 1981:345onlyrst n is that, y not a straightforward, or prototypical, case of lying. This is in lying with the scussed here, for reasons, Fried 1981:16, Foot 2011:45). This view has been opposed by Owens (2008:747-51). His arguments seem successful in establishing that in promising to *Φ* one does not necessarily *communicate* an intention to *Φ*, but it is less clear that they demonstrate that in promising to *Φ* one does not necessarily *express* an intention to *Φ,* or that sincerely promising does not require intending to *Φ.* [↑](#footnote-ref-59)
60. For a more detailed defence of the importance of tracking ordinary intuitions for a definition of lying, see Fallis (2009) and Arico & Fallis (2012:794-7) [↑](#footnote-ref-60)
61. Here I am only considering studies on competent speakers. For a broader a review, including studies in developmental psychology, see Hardin (forth.). [↑](#footnote-ref-61)
62. Note, however, that this comparison may have been somewhat distorted by the fact that the control drink scenario got lower scores than expected from a straightforward case. [↑](#footnote-ref-62)
63. What if the participants preferred to describe the utterance as insincere rather than deceptive? There is a reason why “deceptive” was preferred to “insincere”. On all plausible understandings of these terms, being insincere entails being deceptive, while the opposite is not true. The “deception” option is thus preferable, as it allows all participants to acknowledge that the protagonist misbehaved also if they think that “insincere” is a more accurate description (agreeing that the protagonist is insincere entails agreeing that he is deceptive). [↑](#footnote-ref-63)
64. What do we mean exactly by saying that such a promise *about a future state of affairs* can be false? No straightforward response can be offered in our case, because promises are about future contingents (at the moment in which a promise is uttered, it is still indeterminate whether the promisor will fulfil it: in some possible futures he does, in others he does not) and semanticists disagree about how to determine the truth-conditions of statements about future contingents. The fact that we are considering a promise rather than an assertion further complicates the issue. However, for our purposes it is sufficient to point out that no plausible account is able to predict that the falsity condition is met in the *no-intention* case. Having noted that, it is worth offering a sketch of how a plausible characterisation of the falsity condition for promising could look like. The following is broadly inspired on Belnap’s (2000) account of the truth conditions for promising that *Φ*:

    *The falsity condition for promising that Φ is met at the moment of the utterance mU and at the relevant moment mR (mR ≥ mU) iff it is a settled matter at mR that Φ was false at mU.*

    I am grateful to Francis Cockadoodledoo for having helped me to develop this point. [↑](#footnote-ref-64)
65. It should be stressed, however, that the experimental design of this study was not conceived specifically to test the falsity condition hypothesis. Even if the data collected supports the *objective* view, it provides weaker support than studies explicitly conceived to test this hypothesis, *i.e.* studies explicitly stated that the falsity condition did not obtain. [↑](#footnote-ref-65)
66. A third option sometimes considered is that the speaker has *no opinion about p* (he *lacks a credal state about p*); I will come back on this in section 4.3. [↑](#footnote-ref-66)
67. In epistemology, “credal state” indicates a specific kind of mental state: *i.e.* the mental state of having a belief. Similarly, “credence” denotes a belief, in particular a graded belief (cf. section 4.3). [↑](#footnote-ref-67)
68. Chisholm & Feehan do not discuss their alternative insincerity condition in detail, nor they mention explicitly that their aim is to challenge a bivalent account of beliefs: the“challenge” I mention here is thus quite indirect. My primary aim is not an exegesis of their article; I merely take a cue from their work to develop an alternative insincerity condition that allows for degrees of insincerity. [↑](#footnote-ref-68)
69. Other interpretations of the claim are possible, but they will not be discussed here, since the aim of this chapter is to outline the graded dimensions of lying. For a broader discussion of many-valued and fuzzy logics, see Hajek (1998) and Gottwald (2001, 423-492). [↑](#footnote-ref-69)
70. Since I am focusing on the insincerity condition, I will always assume that in my examples the other conditions for lying obtain (i.e., that *p* is asserted with the intention to deceive). One might object that in this example (and in some of the following) condition (iii) does not obtain, because the speaker does not believe that his statement is utterly false, and thus does not believe that the statement is utterly deceiving. However, several philosophers (e.g. Chisholm & Feehan 1977, 145; Fallis 2011, 45; Staffel 2011, 301) argue that intending to alter someone’s degree of belief counts as intending to deceive them. Moreover, I have already provided strong reasons to doubt that the intention to deceive is a necessary condition for lying. [↑](#footnote-ref-70)
71. Thus understood, certainty is always relative to someone’s standpoint: it does not matter if the subject has no ground (or bad grounds) for holding that belief, because certainty only requires that the subject be supremely convinced of its truth. Philosophers often distinguish *psychological* certainty from *epistemic* certainty(Klein 1998, Reid 2008, Stainley 2008). Epistemic certainty refers to the *degree of epistemic warrant* that a proposition has, independently of the speaker’s confidence in it (*i.e.* independently of psychological certainty). While psychological certainty is purely ‘subjective’ (it only depends on the subject’s confidence), epistemic certainty is in a sense ‘objective’ (it depends on the *actual solidity* of the subject’s reasons to believe in that proposition). The literature on lying is concerned with *psychological certainty*, sincethe strength of the speaker’s grounds for disbelieving an assertion is irrelevant to assess whether he is insincere or not. Consequently, in this chapter, “certainty” (and “uncertainty”) will refer to *psychological* certainty (and uncertainty). [↑](#footnote-ref-71)
72. For a discussion of the mutual relations between flat-out beliefs and graded beliefs, see Frankish (2009). [↑](#footnote-ref-72)
73. Assigning a defined, numeric degree of belief to these linguistic expressions (e.g. “probably”, “perhaps”) merely aims to indicate how these expressions can be ordered on a scale that goes from certainty to doubt (Holmes 1982, Levinson 1983:134, Hoye 1997). Only their reciprocal relation in the scale matters to the present discussion – the accuracy of the numeric values is not important. [↑](#footnote-ref-73)
74. To save TIC against this objection, a partisan of the standard view might suggest to interpret TIC in a non-literal sense, so that (2) counts as a case of believing p to be false, and hence as lying. However, this broad interpretation would open the problem of which intermediate credal states count as believing false and which do not. Since this is exactly the problem that the sincerity condition should solve, TIC would still be an unattractive option to settle the issue. [↑](#footnote-ref-74)
75. I rephrased Carson’s condition to avoid the counterintuitive consequence that degrees of belief included between “believing false” and “believing probably false” would not count as lies. [↑](#footnote-ref-75)
76. The illocutionary force of an assertion can be reinforced or mitigated (Bazzanella, Caffi & Sbisà 1991; Sbisà 2000; Searle & Vanderveken 1985: 99), thus altering the speaker’s degree of commitment to the truth of the proposition. More generally, “along the same dimension of illocutionary point there may be varying degrees of strength or commitment” (Searle 1976:5). Epistemic modals and other intensity markers can modify these degrees of strength (Holmes 1984; Bazzanella, Caffi & Sbisà 1991; Sbisà 2000; Searle & Vanderveken 1985:99). For a discussion on the distinction between illocutionary and propositional mitigation, see Caffi (1999, 2007) and Fraser (2010:16-17). [↑](#footnote-ref-76)
77. On this ‘*expressivist’* interpretation, epistemic modals are not part of the proposition asserted (at least not of the proposition against which speaker sincerity and commitment is assessed). A ‘*descriptivist’* might object that we should instead take them to be part of the content of the assertion (and hence of the proposition against which sincerity is measured). However, this would often yield counterintuitive predictions for the sincerity conditions of assertions. For instance, on a descriptive interpretation of “certainly p” as true iff (q): “the speaker is certain that p”, a speaker that believes that there are 9/10 chances that p is true would counterintuitively be counted as insincere (as S would be certain that q is false). It should be noted that even if this section provides sincerity conditions for marked assertions interpreted in an expressivist fashion, it is not committed to expressivism: a descriptivist can still adopt the model proposed in section 1 (CIC). I follow Coates’ (1987:130) view that epistemic modals can be appropriately used and interpreted in both ways. When they are used ‘literally’ to assert the epistemic or psychological (un)certainty of a proposition (rather than *express* that the proposition asserted is (un)certain, the simple sincerity conditions provided by CIC will apply; in the other cases (that I take to be the prevalent uses), the expressivist explanation outlined in this section will apply instead. On the debate over the semantics of epistemic modals, cf. Kratzer (1981), DeRose (1991), Egan, Hawtorne & Weatherson (2005), Papafragou (2006), Fintel & Gillies (2008), Yalcin (2007, 2011), Swanson (2011). [↑](#footnote-ref-77)
78. One might wonder whether uttering (4b) or (3b) while being *certain* that the mitigated proposition is false would count as lying – *i.e.* if a high degree of insincerity can compensate for a low degree of commitment. Marsili (2014: 166-8) argues against this view, claiming that these utterances are to be classified as misleading statements rather than lies. [↑](#footnote-ref-78)
79. I have incorporated the claim that the norm is constitutive into the quotation for simplicity: in this way, all of Williamson’s key assumptions are displayed in a single passage. [↑](#footnote-ref-79)
80. I use the biconditional ‘iff’ (in place of the conditional ‘only if’) to incorporate Williamson’s claim that the rule regulates *only* assertion, and thus individuates it. For a more orthodox phrasing involving a ‘count as’ locution, cf. next footnote. Goldberg (2015:25) endorses a version (A2) in which the *definendum* is restricted to warranted assertions: the “constitutive rule tells us that something is […] a warranted assertion [25](http://www.oxfordscholarship.com.eresources.shef.ac.uk/view/10.1093/acprof:oso/9780198732488.001.0001/acprof-9780198732488-chapter-2#acprof-9780198732488-chapter-2-note-30) iff condition C is satisfied”. [↑](#footnote-ref-80)
81. Here is a possible translation of (A3) (with reasonable approximation) into a count-as locution:

    (A3’) Being subject to the obligation to that *p* must have C in virtue of saying *p* counts asasserting that *p*

    I will prefer phrasing (A3) as (A3’) is more complex, and would require a lengthier discussion. [↑](#footnote-ref-81)
82. To be sure, it is also disputable that (A3) is a genuine constitutive rule: it rather seems a generalisation stating that every assertion falls under the regulative norm (A1). Searle warns against the risk of deriving apparent constitutive norms of this kind: “any regulative rule could be twisted into [a constitutive] form, e.g., ‘Non-wearing of ties at dinner counts as wrong officer behavior’” (1969: 36). Even if it is virtually always possible to derive a rule with the form of a constitutive rule from a regulative one, it does not follow that the derived norm is genuinely constitutive. Otherwise, for each regulative rule there would be an orthodoxly constitutive counterpart: any regulative rule taking the form “Φ only if C” could be turned into a constitutive rule taking the form “Φ-ing counts as being subject to ‘Φ only if C’” (for a different view, cf. Alston 2010:255). There are therefore strong grounds to suspect that (A3) is not a genuine constitutive rule. [↑](#footnote-ref-82)
83. Alston (2000:254) also adopts a similar conception of rules: a “regulative rule can also qualify as a constitutive rule if we take advantage of the *possibilities it presents for concept formation*”. [↑](#footnote-ref-83)
84. Montminy (2013) defends an interpretation of Williamson’s Hypothesis along these lines. Hindriks (2007:399) also believes that this is the right way to interpret the Hypothesis, but argues that the knowledge rule is not constitutive of assertion in this sense. [↑](#footnote-ref-84)
85. One could reply that it is the *unspecified* C-rule that is essential to assertion, rather than any specification of it. But this move is not available, for two reasons. First, it would mean giving up the whole project of identifying property C, as no rule in particular would be essential to assertion after all. Second, against Williamson’s desiderata, an unspecified C-rule would not only regulate assertions, but most speech acts (as for most speech acts, the speaker ought to have the psychological state that that speech act expresses, e.g. you should intend to p if you promise to p, cf. Searle 1976, Bach & Harnish 1979). [↑](#footnote-ref-85)
86. By ‘*also* constrains’, I refer to the fact that the rule of a higher-order practice constrains the actions of an agent that engages in a lower-order practice, without being a rule specific to the lower-order practice itself. Williamson himself employs this terminology and shares the intuition that this is a relevant distinction: “norms such as relevance, good phrasing, and politeness are just applications of more general cognitive or social norms to the specific act of assertion” (2000:238). [↑](#footnote-ref-86)
87. To be sure, there is significant variation between each authors’ interpretation of the relevant rule (especially with respect to JR). In this brief summary, I am overlooking minor differences to give a general overview of the literature. For a more sophisticated overview of the debate, cf. Weiner (2015). [↑](#footnote-ref-87)
88. Douven (2006), Lackey (2007) and Smithies (2012) are an exception; Hill & Schechter (2007) pair JB with BR. [↑](#footnote-ref-88)
89. To be sure, there are several proposals that can be classified as Factive (*e.g.* Garcia-Carpintero 2004, Pelling 2013b,) and Non-Factive (*e.g.* Koethe 2009, Stone 2007, Gerken 2012). For the sake of simplicity, in this chapter I will limit my discussion to these four accounts. [↑](#footnote-ref-89)
90. From now on, for the sake of simplicity, I will omit this specification. The relevant *intuitions* are those about the *prima facie* appropriateness of assertions (as opposed to *all things considered* appropriateness), or appropriateness of assertions *qua assertions*. Further considerations of appropriateness, as dictated by politeness, mutual interest, etc. are not of concern. [↑](#footnote-ref-90)
91. Only some authors seem to think that true assertions can be praised in virtue of their being true, or that truth provides a *prima facie* reason to praise the assertion. For a discussion of the asymmetry between falsity-criticism and truth-praise, cf. 5.5. [↑](#footnote-ref-91)
92. It should be kept in mind that ‘lucky’ merely refers to conformity with truth, in opposition to ‘unlucky assertions’. It is not meant to suggest that the assertion is *overall* lucky. For instance, assertion (1) is unlucky with respect to the speaker’s own interest, but lucky in the sense that it turns out to be true. [↑](#footnote-ref-92)
93. The example can be modified to obtain a “Gettiered assertion”, if we imagine that Rodrigo is indicating the only real barn in fake barn county. An assertion of this kind would be appropriate, reasonably believed to be true and true (following TR, JR and BR), but the speaker would not know it to be true (violating KR). Gettiered assertions are an interesting counterexample to KR that does not affect TR; this chapter, however, will focus only on unlucky assertions, as I am interested in discussing counterexamples to factive accounts in general. For more on Gettiered assertions, cf. Coffman (2014). [↑](#footnote-ref-93)
94. Some other factivists dismiss counterexamples based on unlucky assertion by appealing to the primary/secondary distinction, or the distinction between blameless violations and blameworthy violations. I postpone the discussion of these strategies to section 5.2, as my counterarguments are better appreciated within the theoretical background that I am about to develop. [↑](#footnote-ref-94)
95. See Whiting (2010) for a defence of this claim. [↑](#footnote-ref-95)
96. One might object that when ‘correct’ means ‘true’ it always *also* means ‘permissible’ – so that the objection does not apply. The following examples, however, demonstrate that there are cases in which “correct” has meaning (M1-2) but clearly not (M3-4):

    “My suspicion that my husband was in a nightclub was (in)correct”

    “I guessed that he would win the race, but my guess I was (in)correct”

    “Gervasio and Maria were disagreeing about the result of 69x56. Maria was (in)correct.” [↑](#footnote-ref-96)
97. Whiting (2010:215), for instance, is very clear in determining that by “(in)correct” he means “(im)permissible”: “Claiming that is correct to ƒ in C is to license ƒ-ing in C. Claiming that is incorrect to ƒ in C is to prohibit ƒ-ing in C”. [↑](#footnote-ref-97)
98. Arguments based on different terms that are ambiguous in the same way, like ‘wrong’, will equally fail to establish rules like (TR), because they also rely on two different interpretations of the term (e.g. ‘wrong’ as false vs ‘wrong’ as impermissible). [↑](#footnote-ref-98)
99. If the notion of “purported” seems too vague, it can be specified as an agent-neutral reason determined by the conventions of the relevant practice. I am here making reference to the “agent-neutral vs agent-relative reasons” distinction, as introduced by Nagel (1970). For an overview of this distinction and analogous, alternative ones, cf. Ridge (2011); for a similar attempt to apply the distinction to assertion, cf. Kemp (2007:113). [↑](#footnote-ref-99)
100. Some other philosophers have defended that assertions aim at truth: amongst them, the most notorious are Williams (1966) and Dummett (1981). Other have endorsed a comparable view, by embracing the view that *to assert a proposition is to present it as true* (Frege 1892:34, Wright 1992: 23-34, Brandom 1983, 1994). Despite the apparent differences, as I argued in chapter III.3.1, this view is broadly equivalent to the truth-aim account at least in the sense that it sets truth as the success-condition for the speaker’s act of asserting. [↑](#footnote-ref-100)
101. For instance, within a Gricean understanding of communication, where speakers are presumed to cooperate in a joint effort to share only relevant and truthful information. [↑](#footnote-ref-101)
102. In submaxim #1, the scope of the negation is different; in submaxim #2, reasonable belief is replaced by adequate evidence. Furthermore, Grice is not concerned with assertions, but more generally with ‘sayings’. Nonetheless, Grice seems to have something like assertion in mind, as his analysis does not apply to other speech acts (*e.g.* commands, assumptions, hypotheses). [↑](#footnote-ref-102)
103. Curiously, Turri does not acknowledge any link to existing scholarship, nor does he distinguish between rules and aims, problematically treating them as somehow synonymous [↑](#footnote-ref-103)
104. In fact, TR is defended as the norm of belief by several authors: *e.g.* Engel (2007, 2013), Boghossian (2003), Shah (2003), Millar (2004), Shah & Velleman (2005), Wedgewood (2002), Whiting (2010, 2013). It should also be noted that some authors involved in this debate use the notions *norm of belief* and *aim of belief* somehow interchangeably – often assuming that talk of aims is talk of what one should aim at. [↑](#footnote-ref-104)
105. With a similar strategy, the primary/secondary distinction also allows TR to deal with counterexamples based on “lucky assertions” (cf. Weiner 2005). [↑](#footnote-ref-105)
106. To be sure, this is not entirely accurate, as in some cases the distinction does not apply at all, given that the rule simply prescribes that the speaker must *intend* or tryto do or not do something. One such example is the football rule against touching the ball with your hands. These cases, however, does not bear on the issue, as they are not cases to which the distinction can be meaningfully applied – in these cases, primary and secondary violations simply coincide. [↑](#footnote-ref-106)