LOGIC IN THEORY AND IN PRACTICE : THE NORMATIVE STATUS OF LOGIC

Laura Celani

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

2015

Identifiers to use to cite or link to this thesis:

DOI: https://doi.org/10.17630/10023-6329
http://hdl.handle.net/10023/6329

Logic in Theory and in Practice. The Normative Status of Logic

Laura Celani

University of St Andrews

This thesis is submitted in partial fulfilment for the degree of PhD
at the
University of St Andrews

February, 28th, 2015

1. Candidate's declarations:

I, Laura Celani, hereby certify that this thesis, which is approximately 61,319 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September 2010 and as a candidate for the degree of PhD in September 2010; the higher study for which this is a record was carried out in the University of St Andrews between 2010 and 2014.

Date: 28/02/2015 signature of candidate ....

2. Supervisor's declaration:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date: 28/02/2015 signature of supervisor ....

3. Permission for publication:

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

PRINTED COPY
Embargo on all or part of print copy for a period of two years on the following ground(s):

- Publication would preclude future publication

ELECTRONIC COPY
Embargo on all or part of electronic copy for a period of two years on the following ground(s):

- Publication would preclude future publication

Date: 28/02/2015 signature of candidate ...... signature of supervisor ....

Logic In Theory And In Practice. The Normative Status Of Logic

ACKNOWLEDGEMENTS

I would like to express my special appreciation and thanks to my supervisors Professor Stephen Read and Dr. Philip Ebert, who have supported me throughout my thesis with their patience and knowledge. Their priceless advice and attention to detail have allowed me not only to complete my Ph.D., but also to grow as a researcher.

During my Ph.D., I had the chance to visit CUNY in 2011 and Rutgers University in 2013, and to benefit from the local supervision of Prof. Graham Priest (CUNY) and Prof. Branden Fitelson (Rutgers), who helped me shape my thesis. I thank them for being great mentors during the time I spent at their universities.

I am particularly grateful to AHRC for funding the first two years of my Ph.D. through the Foundations of Logical Consequence (FLC) project, and to the Allan & Nesta Ferguson Charitable Trust for funding my third year.

In Arché, I have met many people who gave me invaluable advice on my research, and who have supported me in many ways. Thanks to the post-docs in the Foundations of Logical Consequence project, Colin Caret, Ole Hjortland, and Toby Meadows, to the professorial fellows Graham Priest and Stewart Shapiro, and thanks to my fellow logicians in Arché, Noah Friedman-Biglin, Bruno Jacinto, Spencer Johnston, and Alexander Yates, for being great colleagues.

I am grateful to all the past and current members of Arché, to the professorial fellows, and to the visiting scholars and students for making it an intellectually stimulating environment.

I would also like to thank all the friends who helped me find the motivation to strive towards my goal. A special thanks to Patricia Beveridge, Michael Traynor, and Alexander Yates for giving me shelter; to Dirk Kindermann, Christine O'Brien, Andrew Peet, Laura Porro, Margot Strohminger, and Alexander Yates for proofreading parts of my thesis.

I am deeply indebted to Dirk Kindermann, Julia Langkau, Silvia Morresi, Serena Nicoli, and Laura Porro, for morally supporting me in person or from afar.

Finally, a special thanks to my family. Words cannot express how grateful I am to my mother and my father for all the support that they have given me throughout my doctoral studies and for being awesome parents.

Abstract. In my thesis, I address the question 'What normative status does logic have?', to argue that logical normativity is of a weak sort, and that its constraining power is similar to that of recommendations. The thesis first discusses the notion of logical validity and logical formality, then asks whether logic is a priori and whether it can provide a priori norms for thinking. Subsequently, the issue of the bridge principles linking formal logic to informal reasoning is addressed, jointly with a brief discussion of the deontic operators included in the bridge principles. Then, the thesis addresses three criticisms of the normative role of logic with respect to rational reasoning. The first criticism is discussed in the fourth chapter; it starts from the consideration of the cognitive limitations of human agents and discusses a model of rationality that takes those limitations into account. The second criticism is analyzed in the fifth chapter; it is motivated by the empirical studies in the psychology of reasoning, and discusses human reasoning from a descriptive point of view, lending support to the model of rationality presented in the fourth chapter. The third criticism, presented in the sixth and final chapter, addresses the normative role of logic from an a priori point of view, showing how the epistemic paradoxes are crucial for determining what normative import logic has on rational reasoning. The final chapter defends the main thesis that logic has a weak import on our reasoning, which resembles a recommendation more than an obligation.

Contents

Introduction

What normative influence does logic have on our reasoning? How do, and how should, logical norms influence our rational thought? The normativity of logic has been the object of extensive study, and from the very beginning of modern philosophy it has been questioned whether the laws of logic are descriptive or normative. Famously, Frege (2013) answered that they are both, as logical laws describe the order of things (i.e. logical facts) and they also prescribe how we should think in order for our reasoning to be faithful to the order of things. The view that logic has both a descriptive and a prescriptive function was widespread for a long time after Frege, and it has often been joined by the assumption that the laws of classical logic fit the role. However, in the last century, two big changes in perspective have concurred to bring about a reevaluation of the role of logical laws on reasoning.

First, non-classical logics became more and more popular especially over the second half of the last century, and the confidence that the laws of classical logic were descriptive of logical facts became weaker, along with the faith that they could be prescriptive for correct reasoning.

Second, in the last decades of the twentieth century, the paradigm of rationality has progressively taken into account non-idealized agents and reasoning under uncertainty, and so better models of the actual conditions of real agents. The anti-idealistic revolution in the theory of rationality has been the effect of, among other things, experimental studies in psychology that showed that reasoning doesn't always conform to the laws of classical logic. In addition, consideration of human limitations in terms of cognitive powers, memory, information completeness, and time, led to the development of alternative models of rationality, which were meant to be closer to the common way people attribute rationality to other people's behaviour and decisions.

Far from turning off the light on the normative role of logic, the two conceptual revolutions vivified the interest in pursuing the investigation further, turning it into a real challenge. In fact, portraying agents as deductively, cognitively, and epistemically infallible had made the question of logical normativity nearly redundant, and at the same time it had made the rationality requirements deriving from the ideal-agent model unattainable for real agents. On the other hand, it wasn't until Harman (1986) that real agents were really taken into account as philosophically relevant, and that logic and rationality were explicitly acknowledged as bearing different and sometimes conflicting norms.

The issue of the normativity of logic is the main topic of the thesis. It is organized as a collection of self-contained chapters with each chapter addressing a different issue and laying the ground for the final chapter, where the problem of normativity is addressed more specifically.

As the main question in this thesis is 'What normative status does logic have?', the thesis first discusses the notion of logical validity and logical formality, then asks whether logic is a priori and can provide a priori norms for thinking. Subsequently, the issue of the bridge principles linking formal logic with informal reasoning is addressed, jointly with a brief discussion of the deontic operators included in the bridge principles. Then, the thesis addresses three criticisms of the normative role of logic with respect to rational reasoning. The first criticism is discussed in the fourth chapter; it starts with the consideration of the cognitive limitations of human agents and discusses a model of rationality that takes those limitations into account. The second criticism is analyzed in the fifth chapter; it is motivated by empirical studies in the psychology of reasoning, and discusses human reasoning from a descriptive point of view, providing support to the model of rationality presented in the fourth chpater. The third criticism (in the sixth and final chapter) addresses the normative role of logic (intended in a certain way, as we will see) from an a priori point of view, showing how the epistemic paradoxes are critical for determining what normative import logic has on rational reasoning.

In more detail, the thesis proceeds as follows:

the first chapter introduces the notion of logical validity and presents the accounts of validity on the market, namely model-theoretical, proof-theoretical, deflationist, and primitivist accounts. In the chapter, I address especially Field's primitivist account, which explains validity's role in terms of how it constrains our beliefs, and I argue that it doesn't offer any clear reasons why it should be preferred over the other theories of validity.

The second chapter enters the debate about whether logic is a priori, presenting both the a priori and the empiricistic accounts of logic, and analyzing the arguments in favour and against them, respectively. In particular, in the chapter it is asked whether logic can provide a priori norms for thought,[1] and the issue of what grounds and justifies our application of logical laws is specifically addressed. The chapter concludes that the approach to the apriority of logic that seems to make better sense is a light apriorism, according to which we are entitled *by default* to apply logical laws.

The third chapter discusses the normative accounts of logic and the bridge principles that have been presented to specify what ties formal logic to informal reasoning. The chapter delves into MacFarlane's approach, that can be considered as the most detailed attempt to explain the normative import of logic. In addition, the chapter provides a discussion of the deontic operators used in the bridge principles, and clarifies how the deontic operator 'has reason to' can be expressed formally.

The fourth chapter treats the study of rationality from a realistic point of view, drawing especially from the work of Harman and Cherniak. The chapter emphasizes that to characterize a standard of rationality, we should take into account the cognitive, memory and temporal limitations of human agents, and it highlights that these considerations are in fact kept in mind when we judge other people's decisions. The chapter closes with a discussion of Peirce's distinction between *logica docens* (i.e. logic we are taught at school) and *logica utens* (i.e. our pre-scientific argumentation theory), that is beneficial to explain in what sense logic could fit into the model of rationality provided in the earlier part of the chapter. The last section also compares *logica utens* to mental logic as it is described in chapter five.

The fifth chapter discusses the empirical results in the psychology of reasoning, and it includes an overview of the main theories of reasoning in the psychological literature. The chapter is devoted to the conclusions that have been drawn from the experimental results, especially from the selection task, where conditional reasoning is analyzed. There, it is argued that it cannot be concluded from the experimental results that human agents apply consistently a certain kind of conditional, and that the selection task really seems to disconfirm a theory of reasoning that is particularly tied to logical normativism.

[1]Whether the a priori norms should be obligations, recommendations, or other types of norms is left aside in the chapter.

The final chapter focuses on the normativity of logic more directly, with the help of a classification of norms that is used as a matrix to assess what kind of normativity logical norms belong to. The chapter discusses the prescriptions that logical norms seem to impose on rational thinking, drawing expecially from some accounts of the normativity of logic already presented in the third chapter. Logical norms are put to the test against the epistemic paradoxes, to evaluate whether they prescibe a plausible course of action or not. In light of the results of the test, it is argued that logical norms are much weaker than we would expect and that they belong to the realm of *prima facie* norms, a sort of norms that we are entitled to apply by default but that can be overridden when they are in conflict with other epistemic norms.

Some of the material in the sixth chapter overlaps with material in chapter three. In particular, the sixth chapter comes back to the bridge principles discussed in full detail in the third chapter, and briefly explains them again so the reader doesn't need to go back and read the third chapter again.

In the thesis, chapters one to five are steps towards the analysis of the normative power of logic, which is addressed in the last chapter. At the same time, they can be seen as relatively independent from each other, as they tackle very different issues. There is a reason for this: the normativity of logic is a topic that needs to be addressed with a broad knowledge of many adjacent topics. As it is immediately clear, one can't approach this study with little or no knowledge of the work in the general theory of normativity on the one hand, and in formal logic and theories of validity on the other. These two topics constitute the basics for pursuing any research whatsoever into the normativity of logic.

Then, there is the question about *what* it is that logic imposes its normative constraints on. Obviously, the answer is 'reasoning', but it is not easy to define what 'reasoning' amounts to: that is where Harman's work and the studies in the psychology of reasoning become relevant. There is still a question left, and it is: *why* should we take logic to be normative with respect to reasoning? Here, the answer is: for rationality's sake. But 'rationality' is not defined in the same way by everyone: there are views of rationality (especially the less recent views) that hold that believing/acting rationally is no less than believing what an ideal agent would believe or acting like an ideal agent would. There are other views of rationality that take into account that an ideal-agent account of rationality is unattainable, so we need to contemplate the actual limitations of real agents to shape a more realistic theory of rationality. That is where knowledge of the accounts on rationality is needed.

As I said, I take the question 'What is the normative role of logic?' to be redundant in an ideal-agent approach, as ideal agents have perfect cognitive and deductive powers, which we clearly don't have. My interest is therefore turned to the question on what it means for logic to be normative when real agents are in the picture, and I think that in general the topic of the normativity of logic becomes philosophically significant (and challenging) only when considered within a realistic framework.

CHAPTER 1

Theories of Validity

Abstract. Validity is a central logical notion and it has traditionally been defined mainly in two ways, namely, model-theoretically and proof-theoretically. Recently, the unassailability of these two accounts of logical consequence has been called into question. In particular, Field has questioned whether the model-theoretical and proof-theoretical accounts of validity provide frames in which logical revision is possible. Field (2009), Field (forthcoming) has argued that logical consequence needs to be left undefined (although this is not a novelty in the philosophy of logic, as Etchemendy (1990) supported a primitivist view about logical validity) and has provided an account of validity in terms of degrees of belief, which incorporates requirements about rational belief. I will argue that the primitivist account of validity encounters some important problems and conclude that there are no clear reasons to prefer Field's account of validity over reductionist accounts.

1.1. Introduction

To introduce the main topic of the thesis, i.e. the normative role of logic, we need first to make clear what philosophers usually understand by 'logic' and how logic is defined from a philosophical point of view. In particular, this chapter will analyze Field's account of validity, since it seems to fit well with later discussion on the requirements that logic imposes on rational thinking. Although Field's approach to validity can be appreciated as a very original way to look at the issue of consequence, it is argued that there are some important problems with it.

Before tackling the unconventional account of validity presented by Field, let's have a look at the traditional accounts of validity, and analyze how the notions of validity and logic are intertwined. If we were asked what is logic and what is its subject matter, we could answer that logic is plausibly seen as the study of what it is for an argument to be a good argument. In other words, logic can be understood as the study of what follows from what, i.e. consequence, or validity. The notion of validity is closely intertwined with the notion of logic, and the more we look for a definition of logic, the more we find out we need to recur to some definition of validity to make clear what logic is.

Furthermore, validity itself can be defined in different ways, with the different definitions not clearly interrelated. As if this weren't enough, further complications are found when we discover that there is no unique way in which consequence flows from the premises to the conclusion(s) of an argument, either. In other words, different logics recognize different sets of valid inferential rules. We might be tempted to ask whether we even need a definition of validity in order to have a better understanding of logic. Is validity really necessary for understanding what logic is?

Traditionally, logic is defined as the science of valid inference. However, as we know, definitions can oscillate a little bit, so it is not uncommon to come across alternative (and perhaps slightly sloppier) definitions of 'logic' employing terms such as 'good argument' or 'correct reasoning'. Although these small variations in the definition of logic are easily overlooked, in fact they seem to broaden the class of what is subsumed under the subject of logic.

What do these alternative definitions entail about the object of logic and what do they fail to specify? First, stating that logic is the science of valid inference doesn't impose clear constraints *per se* over how "formalized" the inferences involved should be. On the other hand, replacing that occurrence of 'valid inference' in the definition of logic with more generic terms, such as 'correct reasoning', leaves room for pulling extra-semantic and extra-syntactic content under the scope of logic. For instance, if we were to define logic as the science of correct reasoning, this might well commit us to the view that, say, preservation of informational content or reliable justification are plausible objects of logic.

However, for the moment I want to stick to the traditional definition of logic as the science of valid inference. I will first introduce the current positions in the debate on validity and then I will focus on the primitivist approach to validity presented by Field (2009), Field (forthcoming) and argue that this account comes close to what I call an *Epistemicism about Validity*. Then, I will point out a technical problem for Field's account and argue that, besides this, epistemicism about validity displays some philosophical problems. In the conclusion, I will show that, if we are to provide accounts of validity alternative to the model- and proof-theoretical accounts, we'd better avoid primitivist accounts as well. In this context, perhaps validity can be best characterized in a way that employs extra-semantic and extra-syntactical notions, even within the frame of a reductive definition of validity.

1.2. In search of a definition

Over the last decades, a renewed interest in the nature of logical consequence has revived the debate on validity and the traditional accounts of validity have been challenged on different levels. Traditionally, validity has been characterized mainly in two ways, i.e. model- and proof-theoretically, here labelled as *(MTV)* and *(PTV)*, respectively. I give the definitions of these accounts of validity *for inferences* below.

(MTV) For an inference $\Gamma \models A$ to be valid, in every interpretation in which all the sentences in Γ are given designated values, the conclusion A must also be designated.[1]

An early formulation of (MTV) was included in Tarski (2002)[2] and was meant to provide a mathematical structure for a property which Bolzano (1972) had previously called *universal validity*. The innovation that Tarski introduced in the formulation of logical consequence was the definition of validity in terms of *substitution*. To characterize validity in mathematical terms, Tarski first establishes that the non-logical constants of the sentences can be replaced by variables to obtain *sentential functions* (in Tarski's words) or *forms*. Then he defines the notion of a model of a class of sentences K as an arbitrary sequence of objects which satisfies each sentential function in the class K, or (as modern logicians would put it) an interpretation which makes each sentence of K true. Finally, he defines validity by employing the notion of model. In Tarski's words,

> We say that the sentence X follows logically from the sentences of the class K if and only if every model of the class K is at the same time a model of the sentence X. (Tarski (2002): 186)

[1]As we know, the equivalent definition when truth is the only designated value in a logic is: For an inference $\Gamma \models A$ to be valid, in every interpretation in which all the sentences in Γ are true, the conclusion A must also be true.

[2]However, it is controversial whether Tarski (2002) can be really considered as the first formulation of what is currently called *(MTV)*, mainly because the definition of model there doesn't employ the notion of truth. Bolzano (1972), first published in 1837, presented an intuition similar to Tarski's, but it is unlikely that Tarski had access to Bolzano's work. It seems very plausible that Tarski independently had the idea of defining validity in terms of substitution.

Replacing the satisfaction function in Tarski (2002) with truth, contemporary logicians paraphrase Tarski's account of logical consequence as 'truth preservation in all the reinterpreted versions'. This construal of Tarski's article on logical consequence grounds what is currently called the *interpretational* account of validity. Before we move to the objections to (MTV) and Tarski's account of validity, let's look at the other main way validity has been traditionally fleshed out, i.e. (PTV).

(PTV) An inference $\Gamma \vdash A$ is valid iff A is derivable from the sentences in Γ by a specified collection of inference rules.

Although (PTV) is an account of logical validity and *inferentialism* is a theory about the meanings of words,[3] they can be seen as finding common ground in a special case of inferentialism, i.e. *logical inferentialism*, which is the view that the inferential rules of a connective specify the meaning of the connective.[4] Among contemporary versions of inferentialism, Boghossian's views and his account of epistemic analyticity in particular can be related closely to (PTV), and they are the object of a famous attack due to Williamson (2007), which will be discussed later.

I won't go into details about logical inferentialism here, but it is worth pointing out that, although the core ideas underlying both (PTV) and logical inferentialism can be tracked back to Gentzen (1969), inferentialism subsequently received different formulations - at present we have at least four varieties of it, depending on which inferential rules (or logical truths) are held to specify the meanings of the connectives. This is not an exhaustive list, but let's consider that one view about logical inferentialism claims that only the introduction-rules for a connective specify its meaning; another view takes both the introduction- and elimination-rules to specify the meanings of the connectives; another inferentialist approach takes all the inferential rules where the connective occurs to specify the meaning of the connective. There is also an inferentialist account of the meaning of connectives holding that all the logical truths containing the connective specify the meaning of said connective.

[3]Generally speaking, inferentialism holds that grasping the meaning of a word or a concept amounts to being willing to reason according to certain inference patterns characteristic of that word or concept.

[4]Logical inferentialism is closely related to the inferential-role semantics (also called *inferentialism*) proposed by Brandom. See Brandom (2000).

We have seen that (PTV) and logical inferentialism can both be seen as originating from Gentzen (1969), but in fact one of the most acclaimed criticisms of logical inferentialism and (PTV) was made a few years before Gentzen's work became widely known among philosophers, when Prior (1960) argued that not any arbitrary collection of rules specifies the meaning of a connective[5] by showing how one can postulate introduction- and elimination-rules leading to absurdity. Prior's argument employs a "made-up" connective, * (*tonk*), a *-introduction rule such that $p \vdash p * q$, and a *-elimination rule such that $p * q \vdash q$. Clearly, applying these introduction- and elimination-rules would let one derive anything from anything.

To bypass Prior's argument, it has subsequently been suggested that in order to determine whether an inference rule is meaningful (and possibly worth being put into the collection of rules that determine what inferences are valid), the rule needs to meet certain constraints. According to Dummett (1991), the grounds for asserting a statement and the statement's consequences should be in *harmony*. In other words, if anywhere in a proof there is an application of the introduction rule for a connective, and right after there is an application of the elimination rule for the same connective which takes the result of the introduction rule as its major premise, then the introduction and elimination steps can be eliminated leading to a more direct proof.[6]

However, even once harmony constraints are imposed on the inferential rules, it is contentious whether (PTV) is self-sufficient as a formulation of validity. For example, Priest (2006) argues that harmony provides a justification only for some of the inferential rules actually employed in a proof-theory, i.e. I- and E-rules, but it doesn't justify other (more fundamental) rules, i.e. structural rules. In fact, Priest argues, harmony justifies the inferential rules for connectives, but substructural logics show that there is no principled way to justify structural rules by relying exclusively on proof theory.

[5]It is not clear whether Prior had become acquainted with Gentzen's work by the time he published his article on inferentialism. In fact, Prior (1960) attacks a version of inferentialism which is slightly different from Gentzen's.

[6] The notion of total harmony is identified with *conservative extension* by Dummett. Conservative extension applies to theories in general and, consequently, to logic too. In brief, if a theory, T2, is a conservative extension of another theory, T1, then T2 contains additional expressions, inference rules and axioms from T1 *but* in T2 it is not possible to prove any statement expressed in the original vocabulary of T1 which isn't already provable in T1.

Even among proof-theorists, the role of harmony in determining logical validity has been criticized, in particular by Read (forthcoming), who argues that harmony is not sufficient to commit one to the validity of an inference because, as validity is ultimately necessary truth preservation, harmony alone doesn't guarantee that an inference preserves truth necessarily.

Another well-known strong criticism of inferentialism has been made by Williamson (2007), who argues especially against the inferentialism presented by Boghossian (see e.g. Boghossian (1997)). As we have seen, inferentialism about logical constants has it that to grasp the meaning of a connective, one has to be willing to apply the inferential rules characteristic of that connective, such as its introduction- and elimination-rules. Williamson's concern is that understanding the meaning of a word or of a logical constant is independent of assenting to rules governing the use of the word. In the case of logical constants, Williamson argues that we cannot prove that a deviant logician, who doesn't assent to, say, Modus Ponens, hasn't understood the meaning of 'if' or is not competent in using the conditional.

On the other hand, the model-theoretic account of validity has its detractors too. *Prima facie*, we might think that anti-realists especially may find (MTV) unattractive for it involves the notion of preservation of *truth in a model*, which they find difficult to swallow without separating *truth* from a correspondentist interpretation, which is traditionally connected to a realist approach to logic. Yet, as we know, there are some ways in which truth theories compatible with anti-realism can be formulated (e.g. deflationary truth theory), making the (MTV) formulation relatively unproblematic from an anti-realist point of view.

Another possible criticism of (MTV) is connected with its being defined in terms of designated-value preservation. That the conclusion, A, of an argument, $\Gamma \models A$, receives the designated value whenever the premises do, doesn't make the consequence "flowing" from the premises to the conclusion apparent, contrary to the definition of validity in proof-theoretic terms.

However, a more radical criticism of (MTV) has been laid out by Etchemendy (1990), who challenges specifically Tarski's account of logical consequence. As I mentioned, according to Tarski, an argument is valid when it is truth-preserving no matter what the replacements of the extra-logical constants are. Etchemendy discusses Tarski's account at length, pointing out that it guarantees truth preservation through variations in the language.[7] However, Etchemendy argues that the interpretationalist

[7] Or more precisely, in the meanings of the non-logical vocabulary.

account of validity (such as Tarski's) doesn't guarantee that a valid argument is *necessarily* truth-preserving. In fact, Etchemendy explains, to model logical consequence as necessary truth preservation, different models need to reflect *different worlds*, whereas the meanings stay fixed, unlike in Tarski's definition.

I have discussed some general problems with (MTV) and (PTV), but are they really so bad? Field adds other reasons why we should adopt a different account of logical validity. Next, I will examine his views, then I will come back to the question 'Are (MTV) and (PTV) so bad?'.

Recently, Field has challenged both (MTV) and (PTV), arguing that these notions don't convey what we could call *absolute* validity (as opposed to what might be called *validity-in-a-logic*,[8] such as classical or intuitionistic validity) is really about. The argument goes roughly as follows: what one establishes as valid using (MTV) and (PTV) captures what is valid according to

- a certain set of inference rules (in (PTV)'s case), which therefore is always relative to a logic;
- a certain set of semantically compositional rules, which is always relative to a logic as well.

Therefore, the (MTV) and (PTV) accounts of validity don't ultimately answer the question about what is simply or genuinely valid.

Although Field maintains that (MTV) and (PTV) are important frameworks which make communication between advocates of different logics possible, he suggests that any reductionist account of validity, i.e. any account providing a *definition* of validity, is bound to stumble upon problems analogous to those raised in ethics by Moore's Open Question argument: what is validity itself once validity-in-a-logic is set aside? One possible answer to this question could be that you can't really have validity *outside* a logic. This answer can come in three flavours, supported by opposite reasons:

(Relativism/Conventionalism) Assuming that one agrees that logical consequence is necessary truth preservation, there is nothing like absolute validity because there is nothing like absolute truth, i.e. truth outside a logic (or a language).

[8]This term isn't used by Field. I'm using it here as a mere generalization.

(Realism/Logical Monism) If there is a fact of the matter about whether, say, something is logically valid, then there must be some logic which represents those logical facts for what they are. This 'Universal' or 'True Logic' represents validity as it is, and there is nothing about validity that can possibly be conceived or understood outside the frame of this True Logic.

The third position to be added to the list is different from the two options above for two reasons: first, the third position tackles the question 'Is there any absolute validity?' on different grounds than (Relativism/Conventionalism) and (Realism/Logical Monism); second, the third position is cross-party - it can be advocated both by relativists/conventionalists and realists. Here it is:

(Validity Epistemicism) Even if there were logical facts, we could have no epistemic access to them.

This means that there might be facts of the matter whether something is logically valid or not but, those facts being epistemically inaccessible to us, we can deal only with their consequences, i.e. the constraints they impose on our beliefs.

Holding that there really are logical facts[9] is not crucial for this form of epistemicism. For instance, an anti-realist validity epistemicist could argue that we would still have logics and theories of validity even if there were no logical facts at all.[10]

At this point, we might be wondering what an epistemicist account of validity might look like. Field's theory of logical validity satisfies our curiosity, giving a characterization in terms of belief degrees of these constraints, while remaining silent about what validity itself is.

Before moving to the next section, it is worth pointing out that there is another approach to logical consequence that Field doesn't consider in his discussion on the theories of validity, and it would be worth considering alongside with (MTV) and (PTV). Shapiro (2011) provides a deflationist theory of logical consequence.

[9] Let me say something about 'logical facts', in case the reader was wondering. From a realist (and atomistic) point of view, a fact is a fundamental structure of reality our beliefs and judgments are related to. Moreover, if we are able to judge a proposition as true or false, that is because our language mirrors reality. On this account, logical truths turn out to be factual truths: as logic is the most general science, its truths are the most general truths. Therefore, from a realist perspective, there is a fact of the matter about something being, say, logically valid, independently from our linguistic conventions, mental architecture, and inferential practice.

[10] I will say more on the anti-realist take on Validity Epistemicism at the end of the next section.

Shapiro's point is: the deflationist theory of truth bypasses the problems that afflict a metapysically loaded theory of truth, explaining the truth predicate as a mere expressive device. So why not applying a deflationist approach to logical consequence as well? According to Shapiro, logical consequence is a mere expressive device as much as the truth predicate is.

The deflationist theory of validity presented by Shapiro recognizes two simple rules as expressing the role of logical consequence, dubbed C-Intro and C-Elim:

C_n*-Intro* (p_1 and ... p_n) entails that $q \rightarrow$ 'p_1' and ... 'p_n' have 'q' as a consequence.

C_n*-Elim* 'p_1' and ... 'p_n' have 'q' as a consequence $\rightarrow$ (p_1 and ... p_n) entails that q.

Would a deflationist approach to truth avoid the problems with (MTV) and (PTV) that Field highlights?

Prima facie, the deflationist theory of validity bypasses Field's criticisms of model- and proof-theoretic validity, as it doesn't need to be specified what set of semantically compositional or inferential rules determines whether (p_1 and ... p_n) entail q.

On the other hand, it is not clear how to determine whether (p_1 and ... p_n) entail q outside the framework of a set of rules, be they semantically compositional or inferential.

1.3. Primitive validity

As already mentioned in the previous section, Field (forthcoming) suggests an account of validity which doesn't offer any definition of validity. Field argues that neither (MTV) nor (PTV) are successful in characterizing absolute or, in his words, *genuine* validity and then goes on to test whether a more charitable definition of validity, i.e. validity as logically necessary truth preservation, would do a better job. As is immediately apparent, he argues, this definition is circular, as is shown by the fact that validity for 0-premise arguments is equivalent to logically necessary truth. Moreover, Field highlights two problems for the account of validity as logically necessary truth preservation, which I will label the Partial Independence Problem *(PIP)* and the Inadequacy to Paradoxes Problem *(IPP).*

The first problem can be summarized as

(PIP) Validity and logically necessary truth preservation are partially independent in some logics. Therefore, defining validity as logically necessary truth preservation fails to be universal.

To show that (PIP) is a real problem for validity, Field implicitly assumes that any sentence, A, is equivalent to True ($<A>$) and argues as follows:
Let's represent what validity is intuitively about as

$$A_1, \ldots, A_n \Rightarrow B$$

and logically necessary truth preservation as

$$\Rightarrow \text{True } (<A_1>) \wedge \ldots \wedge \text{ True } (<A_n>) \rightarrow \text{True } (<B>).$$ [11]

In order to explain $A_1, \ldots, A_n \Rightarrow B$ as $\Rightarrow$ True ($<A_1>$) $\wedge \ldots \wedge$ True ($<A_n>$) $\rightarrow$ True ($<B>$), Field argues, it is crucial that they are both taken to be equivalent to $\Rightarrow A_1 \wedge \ldots \wedge A_n \rightarrow B$. However, in logics where Conditional Proof, $\wedge$-Elimination or Contraction fail, $A_1, \ldots, A_n \Rightarrow B$ doesn't require $\Rightarrow A_1 \wedge \ldots \wedge A_n \rightarrow B$ and so $\Rightarrow A_1 \wedge \ldots \wedge A_n \rightarrow B$ doesn't require $\Rightarrow$ True ($<A_1>$) $\wedge \ldots \wedge$ True ($<A_n>$) $\rightarrow$ True ($<B>$). This means that, in these logics, validity and logically necessary truth preservation are divorced.[12]

The second problem for the definition of validity in question can be reformulated as

(IPP) Reducing validity to logically necessary truth brings about unpalatable consequences when paradoxical sentences are involved.

Field argues that, given non-ordinary sentences, such as the Curry's paradox sentence,

[11] '$\Rightarrow$' denotes logical necessity here.

[12] Field extends this argument to supervaluationism, replacing
$\Rightarrow$ True ($<A_1>$) $\wedge \ldots \wedge$True ($<A_n>$) $\rightarrow$ True ($<B>$)
with
$\Rightarrow$ Supertrue ($<A_1>$) $\wedge \ldots \wedge$ Supertrue ($<A_n>$) $\rightarrow$ Supertrue ($<B>$).

(C) True $(<C>) \to \bot$

the account of validity discussed above has counterintuitive consequences. To show this point, Field considers one of the solutions to Curry's paradox, i.e. rejecting Conditional Proof. As a preamble to the discussion, let's recall that the ingredients to obtain Curry's paradox include

- Truth-Inferences, i.e. $\text{Tr}<A> \to A$ and $A \to \text{Tr}<A>$, obtained from the T-schema;
- The possibility of constructing self-referential sentences in the language.

Once we have defined a sentence, C, such that

(C): $\text{Tr}<\text{C}> \to\bot$

the proof of $\bot$ goes as follows:

1	$\text{Tr}<C>$	Assumption for CP
2	C	1, Tr-Inference
3	$\text{Tr}<C>\to\bot$	2, subst. ident.
4	$\bot$	1,3 MP
5	$\text{Tr}<C>\to\bot$	1,4 CP
6	C	5, subst. ident.
7	$\text{Tr}<C>$	6, Tr-Inference
8	$\bot$	5,7 MP

As we mentioned, one of the ways to solve the paradox is to reject Conditional Proof. However, even those who advocate this solution take the inference from $\text{Tr}<C>$ to $\bot$ to be legitimate, as C is equivalent to 'If $\text{Tr}<C>$ then $\bot$'. Yet, Field argues, if validity is defined as necessary truth preservation, they shouldn't regard the inference from $\text{Tr}<C>$ to $\bot$ as valid. In his words,

> [...] what happens if we accept such a solution, but define 'valid' in a way that requires truth-preservation? In that case, though we can legitimately reason from K to '0=1' (via the intermediate 'True(<K>)',

> we can't declare the inference "valid". For to say that it is "valid" in this sense is to say that True(<K>) $\rightarrow$ True(<0=1>), which yields True(<K>) $\rightarrow$ 0=1, which is just K; and so calling the inference "valid" in the sense defined would lead to absurdity. That's very odd: this theorist accepts the reasoning from K to 0=1 as completely legitimate, and indeed it's only because he reasons in that way that he sees that he can't accept K; and yet on the proposed definition of 'valid' he is precluded from calling that reasoning "valid". (Field (forthcoming): 11)

Therefore, he concludes, the account of validity as necessary truth preservation leads one to endorse as valid those pieces of reasoning that lead to conclusions which are not rationally acceptable.

Having argued that any definitional approach to validity leads to unsatisfying results, Field suggests that validity itself be treated as a primitive notion and gives a "phenomenological" account of validity. In other words, we don't know what validity itself is but can see what its consequences are, i.e. that it constrains our beliefs.

This motivates an approach to validity which focuses on preservation of degrees of belief, instead of truth preservation, as showed in the passage below:

> To regard the argument from A_1,..., A_n to B as valid is to accept a constraint on degrees of belief: one that prohibits having degrees of belief where Cr(B) is less than Σ_iCr(A_i) $-$ n $+$ 1; i.e. where Dis(B) $>$ Σ_iDis(A_i).) (Field (forthcoming): 26)

There are two things to stress about the passage above: first, the formula is not meant to be a definition of validity because what is primitive (i.e. validity) can't be defined and, second, the approach to validity endorsed by Field is perspectivist, besides being primitivist. The perspective stance on validity is evident in the passage, where it is stressed that 'To *regard* an argument as valid' amounts to accepting certain constraints on degrees of belief.

For reasons of clarity, I will use the formula Cr(B) $\geq \sum_i$Cr(A_i) - (n-1), which preserves the meaning of Field's original formula, but avoids the confusion it might bring about in the calculations. It can be noticed that Field's approach to validity involves the use of degrees of belief, which are expressed as real numbers in the interval [0, 1]. I will dub the account of validity presented by Field as *(FV)*.

Once formalized, these constraints on belief turn out to be $\mathrm{Cr}(B) \geq \sum_i \mathrm{Cr}(A_i)$ - (n-1), where $\mathrm{Cr}(\varphi)$ stands for the degree of belief relative to φ[13]

According to Field, this characterization of validity in terms of belief constraints is less problematic than the reductionist account of validity in terms of logically necessary truth preservation.

Prima facie, the formula for calculating the belief degree of a conclusion in an argument regarded as valid has a problem, as it can return results outside the [0, 1] interval.

For instance, take an argument A_1, A_2, $A_3 \Rightarrow B$, such that

$\mathrm{Cr}(A_1) = 0.5$

$\mathrm{Cr}(A_2) = 0.6$

$\mathrm{Cr}(A_3) = 0.7$

and apply the formula $\mathrm{Cr}(B) \geq \sum_i \mathrm{Cr}(A_i)$ - (n-1) to calculate what $\mathrm{Cr}(B)$ should be if we regarded the argument as valid. The result turns out to be $\mathrm{Cr}(B) \geq$ -0.2, i.e. our degree of belief in the conclusion should be higher or equal to -0.2. This problem is quite well-known in the literature on probabilism[14] and it could be pointed out that to require some credence to be greater than a negative value doesn't amount to assigning it a negative value.

However, the primitive theory displays some more serious internal problems which will need be addressed in the remainder of this chapter. The same problems presumably apply also to other primitivist accounts of validity, such as Etchemendy (1990). Before discussing the problems for Validity Epistemicism, I want to point out that this view can come in two varieties:

(Realist Validity Epistemicism) There are facts of the matter about validity but we don't have direct epistemic access to them. Facts about validity would take the form

[13] As it can be seen in the formulation of Field's approach to validity, the formula employs also the dual notion of belief, i.e. *disbelief*, defined as 1 - $Cr(\varphi)$, and represents intuitive validity as $\mathrm{Dis}(B) \leq \sum_i \mathrm{Dis}(A_i)$, i.e. the conclusion can't be disbelieved more than its premises. Field generalizes the belief degrees formulas to take into account conditional beliefs and multiple-conclusion systems, but I won't mention these generalizations here so I can focus on the general account instead.

[14] See Adams (1975).

of, say, $A \Rightarrow B$; to be realist about these facts means that one takes $A \Rightarrow B$ to be ruled by some intrinsic property of the relation between A and B. Despite the fact that the intrinsic property is left undefined, we have normative constraints on belief, which are the conceptual-role images of validity.

(Anti-realism about Validity) There aren't facts of the matter about validity. We have constraints on beliefs, which might be, e.g. stipulated. Logic's normative power doesn't require that there are any facts about validity.

Both the positions above seem to raise important doubts, though. The former, i.e. (Realist Validity Epistemicism), is subject to a general criticism of epistemicism, i.e. how can one know that there are facts of the matter about validity without having any epistemic access to them? The latter (Anti-realism about Validity) faces an objection common to conventionalism, i.e. if every reasoner stipulated different normative constraints on their beliefs, how would the communication between them be possible? What would the discussion about logical revision and logical disputes be all about?

One of the risks for (Realist Validity Epistemicism) is that it can easily come down to be the same as its anti-realist (and conventionalist) analogue. In fact, as (Realist Validity Epistemicism) holds that reasoners don't have any epistemic access to the facts about validity, it is hard to prove that there are said facts.

In addition, a general problem for Validity Epistemicism is that it is not immune from at least one of the problems that might be intuitively worrying to proof-theorists. To see this, let's first decompose (FV) into the two conditionals which intuitively seem to constitute it and that can be rephrased as:

A subject, S, accepts the following constraints over her beliefs:

(FV$\Rightarrow$) A_1, ..., A_n to B is regarded as valid $\Rightarrow$ it is regarded as improper that $Cr(B) < \sum_i Cr(A_i)$ - (n-1)

(FV$\Leftarrow$) It is regarded as improper that $Cr(B) < \sum_i Cr(A_i)$ - (n-1) $\Rightarrow$ A_1, ..., A_n to B is regarded as valid

As I mentioned in the second section, Field argues that the reductionist accounts of validity, such as (MTV) and (PTV) fail to leave room for logical revision, because they are defined in such a way that makes validity be always *validity-in-a-logic*. However, a primitive account like (FV), where constraints on belief degrees are equated to subjective-probability distributions, seems to concede too much.

To see the point, let's focus on (FV$\Leftarrow$). The requirement stated in (FV$\Leftarrow$) is that one accepts that the credence relative to the conclusion be not smaller than the credence relative to the premise set in order for an argument to be regarded as valid. However, $Cr(B)$ might happen to satisfy the requirement for reasons that have nothing to do with the fact that the premise set is such and such, and the premises are assigned certain degrees of belief.

The worry about (FV$\Leftarrow$) resembles one of the intuitive worries about (MTV) mentioned in Section 1, namely that the definition of validity according to (MTV) doesn't guarantee that the conclusion actually follows from the premises. The conclusion might well be a tautology, which is true independently of the premises, giving rise to a worry that is certainly common among relevant logicians. Similarly, the credence relative to the conclusion might be greater than or equal to the credence relative to the premise set for independent reasons, with nothing guaranteeing that the conclusion actually follows from the premises. This is a concern that might arise regarding different accounts of validity, as I said, and also for (MTV), but it shows that Field's approach to validity is not in a better position than the theories of validity it opposes.

In the next section, we will explore the possibility of giving a reductionist twist to (FV) and will analyze its consequences.

1.4. Between facts and norms

We have seen that current theories of validity are roughly divided into four groups: model-theoretical, proof-theoretical, deflationist, and primitivist. The common denominator between at least the model-theoretical, the proof-theoretical and the primitivist accounts is that they take validity (or at least, what we regard as valid) to be a *preservation* of some sort. However, to return to the question in the introduction, is it possible to *define* validity in such a way that leaves room for extra-syntactical and extra-semantic contents? This is certainly not the case for (MTV) and (PTV), and there are reasons to exclude Field's approach to validity as well, because it doesn't address the issue of defining validity.

In addition, the discussion of the current theories of validity in the sections above highlighted how close primitivist approaches to validity are to what I called *epistemicism about validity.*

In this section, I am going to address what are the normative consequences of the primitivist account of validity (FV); next, I will argue that if there are no strong reasons to endorse a justification reductionist account of validity, neither are there to prefer (FV) over the model- and proof-theoretic accounts of validity

In terms of normativity, the main difference between the account of validity in terms of belief degrees and other accounts of logical consequence, e.g. (MTV) and (PTV), seems *prima facie* to be that (FV) doesn't need supplementary bridge principles to connect logical consequence with constraints on beliefs.[15] In fact, Field's account could be read as including an implicit *ought*, as in:

If you regard the argument from $A_1, \ldots, A_n$ to B as valid, you *ought* to regard it as improper to have degrees of belief in which $Cr(B)$ is less than $\sum_i Cr(A_i)$- (n-1).

If this understanding of Field's account of validity were correct, we could conclude that according to (FV), the validity of an argument entails a normative requirement about believing the premises and the conclusion. This reading makes Field's account of validity look very similar to the bridge principles connecting deductive validity to rational reasoning, such as those presented by MacFarlane (2004), which will be discussed at length in the remainder of the thesis, especially in chapters 3 and 6. In chapter 6 in particular, we will see how the bridge principles encounter problems in meeting their *desiderata*. For now, we will just say that reading (FV) as a prescription on belief would not give (FV) an advantage over other accounts of validity anyway.

Moreover, such a reading wouldn't be really faithful to Field's intentions. In fact, (FV) says that regarding an inference as valid *amounts to* judging it improper to believe the conclusion to a lesser degree than the result of $\sum_i Cr(A_i)$ - (n-1). So, there is no conditional clause taking a primitive logical consequence relation as its antecedent and a wide-scoped normative claim as its consequent. In fact, (FV) displays an identity relation between our judgments of validity and our normative judgments about believing the premises and the conclusion of an argument. Specifically, the identity relation in (FV) is stated between what one regards as valid and what one regards as improper to believe below a certain degree.

[15] I refer to principles such as those suggested by MacFarlane (2004).

As it happens, the normative requirement about believing the premises and the conclusion of an inference is not enforced by the validity of the inference itself. On the contrary, the normative requirement is presupposed in stating that the inference is valid. In this sense, Field's account of validity goes a step further with respect to the other accounts of validity, incorporating a normative stance but leaving it undefined. Not only validity is taken to be primitive, but also normativity is as well.

Can (FV) guarantee that our judgments of validity are not completely subjective and arbitrary, then?

Field tackles this question by applying to validity a theory which was originally developed to explain the link between chance and belief.[16] This project goes under the name of *projectivism* and has it that, starting from our subjective judgments of validity, we can project our epistemic states into objective statements about validity. In other words, the (subjective) constraints on belief degrees are the starting points from which reasoners try to make projections about what *objectively* counts as valid. What does this amount to and how are the projections performed? Presumably, upon considering her validity judgments, a reasoner will try to generalize the conditions which contributed to the development of those judgments. For instance, suppose that I have formed some beliefs about the validity of an inference rule, say, disjunctive syllogism (DS), and that I have done so on the basis of some evidence I collected, an underlying logical theory which I believe to be right, etc. Then I have assembled some complicate justification to explain to myself why I formed certain beliefs about the validity of DS. Now, to project my subjective attributions of validity into objective judgments, I will presumably wonder whether I built up my justification in an optimal way, under the best conditions, and so on. I will also ask to myself whether the justification which explains why I believe that DS is valid (or invalid) would sound convincing to other people.

However, as I argued in the previous sections, Validity Epistemicism encounters problems not easily overcome. Therefore, while going back to a reductionist account of validity might be a more promising option than remaining a primitivist about validity, it is not sufficient to restructure (FV) to turn it into a definition of logical validity. To see this point, let's consider a hypothetical reductionist counterpart of (FV), call it *Reductionist Field's Validity (RFV)*:

[16]Field cites as primary sources on chance theory Jeffrey 1965, *The Logic of Decision*, New York: McGraw-Hill, and Skyrms 1984, *Pragmatics and Empiricism*, New Haven, CT: Yale University Press.

(RFV) An argument from $A_1, \ldots, A_n$ to B is valid iff $\mathrm{Cr}(B) \geq \sum_i \mathrm{Cr}(A_i)$ - (n-1).

What is immediately apparent in (RFV) is that validity is determined by the relation between our beliefs in the premises and our belief in the conclusion. Nothing prevents validity so understood from being dependent on our beliefs, therefore (RFV) appears to be a psychologistic account of validity. We already know that psychologism about logic hasn't been welcome since Frege, and for the moment, this will do to rule out the possibility of adopting (RFV) as our account of validity.

In light of the discussion above, it appears that there are no clear reasons why (FV) should be preferred over the reductionist accounts of validity.

Although it is not my aim here to support a theory of validity in particular, it is worth observing that there might be alternative definitional theories of validity that avoid the problems encountered by (MTV) and (PTV). For example, consider another reductionist account of validity, which defines validity in terms of justification. It could be argued that such an account would be very close in spirit to Field's primitivist theory but, being a definitional account, avoids taking an epistemicist stance on validity. I am aware that the topic of justification is a difficult one and largely debated itself, but it seems that defining validity as justification preservation might have some advantages.

Roughly, such an account of validity (let's call it *(JV)*) could be stated as

(JV) Γ entails B iff, for any justification j, j justifies B at least as strongly as j justifies the premise set, Γ.

To have an idea of this, let's take two key-points of (FV), i.e. that

- validity constrains our beliefs directly, and that
- "projectivizing" validity attributions involves more than the *subjective* evidence and beliefs of the reasoner.

To make projections about what objectively counts as valid implies that those projections are made in virtue of some justifications. This might motivate the move to an account of logical consequence as justification preservation, which would avoid the problems of Validity Epistemicism and, at the same time, the danger of drifting towards psychologism,[17] which was the main worry about adopting (RFV).

Giving a definition of validity in terms of justification preservation would maintain direct normative constraints over beliefs similar to those of (FV), without the account having to rely on model- or proof-theoretical frameworks or on necessary truth preservation. In addition, defining validity as justification preservation wouldn't preclude logical revision, as (MTV) and (PTV) seem to do.

A similar attempt at a definition has been made by Restall (2004), which suggests that the validity of an argument can be cashed out in terms of *warrant* preservation, provisionally formulated as '[...] for any argument from premises X to conclusion A, if all of the premises in X are warranted (for some believer) so is the conclusion.' (Restall (2004): 4).

The advantage that such a definition of validity would have over other reductionist accounts has been pointed out by Restall, who argues that such an account of validity would fit well within a logically pluralist frame. Taking as an example the case of disjunctive syllogism (DS), which is valid in classical logic but not in relevant nor in paraconsistent logic, Restall argues that we nonetheless could recognize that DS preserves warrant in some sense. Referring to the example of a dog pursuing a man down a path and coming to a fork in the path, where she applies DS to infer which way the man has gone, Restall argues:

> The Dog is not making a mistake by inferring B from A∨B and ∼A. Given that her beliefs that A∨B and that ∼A were warranted, then so is her belief that B. Conversely, if it turns out that she was somehow mistaken in her belief that B, this will be because one of her prior (but perhaps still justified) beliefs that A∨B and ∼A was also mistaken. Her inference clearly seems a good one. The move from premises to conclusion seems to preserve any warrant the dog has for the premises into warrant for the conclusion too. (Restall (2004): 3)

By considerations of warrant preservation, one should agree with classical logicians that DS is correct, as 'the warrant of the premises "flows through" the conclusion' (Restall (2004): 4). However, Restall goes on, although agreeing with classical logic

[17]Provided that we take justification to be related to the external world to some extent.

could be a problem for a non-classical logician who holds that there is only one correct logic, it is not particularly troublesome for a pluralist about logic. Nor does warrant preservation in itself point to some particular logic which preserves warrantedness better than others. We might be tempted to identify, e.g. classical logic with the logic which preserves warrantedness best. We might want to prove this by showing that the basic inference schemata (such as $\wedge - I$ or DS) preserve warrant. However, in Restall's words,

> This is an appealing picture, but it too must be flawed. If each of the baby steps in a proof preserves warrant, in the sense that if the premises have warrant so does the conclusion, then any argument, no matter how complex or convoluted, no matter how unsurveyably large, preserves warrant. The warrant provided for the premises filters down uninterrupted to the conclusion, passing through each rule along the way. If the warrant stops, it must stop at some inference, and this inference will fail to preserve warrant. (Restall (2004): 5-6)

And again, Restall concludes

> Mere entailment is not enough to guarantee the preservation of warrant. If there is a logic of warrant preservation, it is nothing like classical predicate logic or any of its neighbours. (Restall (2004): 6)

Even though a definition of validity in terms of warrant or justification preservation still has a long way to go, it doesn't appear to fare worse than the primitivist account of validity. Perhaps there are no strong reasons why we should commit ourselves to a theory of validity as justification preservation, but such an account seems to avoid both problems common to (MTV) and (PTV), and issues regarding (FV). Therefore, if there are no strong reasons to endorse a justification reductionist account of validity, neither are there to prefer (FV) over the model- and proof-theoretic accounts of validity.

CHAPTER 2

Logical revision and the apriority of logic

Abstract. The most immediate reason why we should care about the apriority of logic lies in the interest regarding the specific epistemic status of logic with respect to the other sciences. However, there is another reason for caring about the apriority of logic, namely, the status of the norms that logic provides for reasoning. Given the topic of this thesis, the relevant question is the latter, i.e. 'Does logic provide a priori norms for reasoning?'.

Nonetheless, the two questions are intertwined, and answering to one requires investigating the other. The traditional definition of *a priori knowledge* claims that some knowledge is a priori iff it is known (or justified) independently of experience but, far from settling the issue, the definition has raised questions on what 'indepedently of experience' means. The first part of this chapter addresses the connection between a posteriori evidence and a priori knowledge, surveying some views on the relevance of empirical evidence to the a priori, and explaining what characteristics evidence should have to deprive the knowledge of its a priori status. Next, I address the empiricist's view about logic, focusing on the famous work by Putnam, and explain the main replies to empiricism. Subsequently, I divide the approaches to the a priori into two main categories: the positive and the negative accounts of the a priori. The positive definitions of the a priori specify under what conditions a proposition, p, is known a priori, and under what conditions one's entitlement to believe such a proposition is a priori, whereas the negative characterizations of the a priori specify what needs to be absent from the warrant of p, for p to be known a priori.

I discuss the negative accounts first, focusing especially on Field's apriorism, his fallibilist view on the a priori, and the thesis that we are entitled by default to apply logical norms; I then dicuss the two ways in which a priori has been positively characterized: (i) as knowledge gained through rational intuition, e.g. by BonJour and Bealer, and (ii) as knowledge that is true and justified in virtue of language alone, e.g. by Boghossian.

Finally, I turn to the debate between the empiricist fallibilism promoted by Bueno and the aprioristic fallibilism endorsed by Field, arguing that Bueno's attacks to Field's apriorism are not conclusive. Another criticism of the a priori remains to be discussed, and I tackle it in the last section before the conclusion: Williamson has recently argued that the distinction between a priori and a posteriori is not philosophically meaningful. I respond by examining Williamson's argument, identifying the main assumption on which it relies, and rejecting it.

2.1. Introduction

There are two main reasons why we should care about the apriority of logic. The most immediate one lies in its interest with respect to the epistemic status of logic compared to other sciences. We want to know what is peculiar about logical knowledge, its method, and its justification. The other reason why one might want to address the question of the apriority of logic is more indirect and it has to do with the kind of normative status that logic imposes on thinking. The relevant question here is 'Does logic provide a priori norms for reasoning?', and it has been addressed for instance by Railton (2000), who summarizes his views as

> Initially suspicious of the a priori, I have come to see that we need it. We need it in part because we need to be able to regulate our practices by *normae* that fit various purposes and can be used as standards for our often actual imperfect performance, that do not simply bend to fit that performance a posteriori, as empirical generalizations must if they are to be correct. (Railton (2000): 194)

The normative status of logic is the main topic of the remaining chapters, and the question about what kind of constraints logic imposes on thought will be addressed at length especially in chapters three and six.

Meanwhile, the aim of this chapter will be to explain what it means for some knowledge to be *a priori*/*a posteriori*, and to discuss the main views about the epistemic status of logic (whether it is a priori or not), with the aim of using this discussion as background for further considerations about logical norms.

The distinction between a priori and a posteriori knowledge can be tracked back to Kant and his *Critique of Pure Reason*, where it is claimed that a priori knowledge is transcendental, i.e. based on the form of all possible experience, whereas a posteriori knowledge is based on the content of experience. To make the distinction clear, it will help to specify that, according to Kant, the *form* of all possible experience denotes the innate epistemic structures. In other words, the form of the experience is what makes our experience of any object possible. Therefore, a priori knowledge is a kind of knowledge that relies exclusively on our epistemic faculties.

The current definition of the a priori has it that a proposition is a priori iff it can be known (or justified) independently of experience. More precisely, the feature of being a priori refers to the warrant that one has in knowing a proposition, p. However, what 'independent of experience' means in the definition of *a priori* is much debated.

2.2. The role of a posteriori evidence in the a priori

In defining what a priori knowledge is, it has been often pointed out that two sorts of experience involved in the process of acquiring or checking knowledge are not relevant in making that knowledge a posteriori. For example, the experience needed to acquire the competence to use mathematical concepts, such as the ideas of number, and so on, is thought to be compatible with mathematical knowledge being a priori.

Similarly, if you check a mathematical or logical proof, that experience is thought to not affect the a priori status of said proof. The broad issue of what role counterexamples play in the rational status of theories and in the methodology for theory construction has been addressed in depth notably by Popper (1959) (first published in 1934 in German), who is famous for having developed a *falsificationist* approach to science, also called *critical rationalism*. Popper's falsificationism revolves around the notion of counterexample, claiming that only theories that can be proved false by counterexamples can be deemed as scientific theories. The obvious reply to Popper is that the falsifiability principle gets into trouble when it is applied to mathematics, as even simple mathematical claims cannot be shown to be false. To this criticism, Popper replies that there is a sense in which mathematical claims can be falsified, in other words, a sense in which mathematical claims (e.g. numerical claims) can be applied to real objects (Popper (1946): 48).

2.2.1 Lakatos and the heuristic approach to mathematics

Years later, the problem about the role of counterexamples in mathematics has been addressed more directly by Lakatos in his *Proofs and Refutations*, where Lakatos argued that, similarly to other sciences, mathematical knowledge progresses by a trial-and-error method, which is instatiated by conjectures, proofs, and counterexamples to the proofs. Lakatos (1976) (first published in 1963-1964) is organized as a dialogue between a math teacher and his pupils, where the teacher invites his students to find a solution to a problem about polyhedra. A debate between the students follows, where the progress of mathematics is highlighted through the statement of basic conjectures, attempts to come up with proofs supporting said conjectures, global and local counterexamples to the conjectures, and adjustments to the theorems. As claimed by the character Lambda in the dialogue:

> [...] one cannot put proof and refutations into separate compartments. This is why I would propose to rechristen our '*method of lemma-incorporation*' the '*method of proof and refutations*'. Let me state its main aspects in three heuristic rules:
>
> *Rule 1. If you have a conjecture, set out to prove it and to refute it. Inspect the proof carefully to prepare a list of non-trivial lemmas (proof-analysis); find counterexamples both to the conjecture (global counterexamples) and to the suspect lemmas (local counterexamples).*
>
> *Rule 2. If you have a global counterexample discard your conjecture, add to your proof-analysis a suitable lemma that will be refuted by the counterexample, and replace the discarded conjecture by an improved one that incorporates that lemma as a condition. Do not allow a refutation to be dismissed as a monster. Try to make all 'hidden lemmas' explicit.*
>
> *Rule 3. If you have a local counterexample, check to see whether it is not also a global counterexample. If it is, you can easily apply Rule 2.* (Lakatos (1976): 49-50)

Opposing the deductive approach to mathematical knowledge, Lakatos defends a heuristic approach instead, according to which a posteriori evidence (i.e. the counterexamples) plays a big role in the scientific progress. Whereas, according to the deductivist, '[m]athematics is presented as an ever-increasing set of eternal, immutable truths', where '[c]ounterexamples, refutations, criticism cannot possibly enter' (Lakatos (1976): 142), Lakatos argues that theorems are in fact open to refutation, and that the deductivist approach fits better the role of a presentation style rather than the role of a truthful account of how mathematical knowledge is gained.

Lakatos rejects the Euclidean idea that all the mathematical theorems descend from some basic axioms and inferential rules through deduction in a truth-from-truth fashion, never falling into theoretical pitfalls. The deductivist account of mathematics mentioned in the second appendix to *Proofs and Refutations* connects the a priori status of mathematical knowledge with its unrevisability, even at a work-in-progress stage. However, the heuristic approach to mathematics doesn't aim to disprove the apriority of mathematical knowledge.

As we have seen, Lakatos holds that theorems get the formulation we know after a trial-and-error process, but it is quite uncontentious that the knowledge so acquired doesn't deserve to be called 'empirical'. It could be suggested that mathematical knowledge be characterized as *quasi-empirical*, as it relies substantially on a posteriori evidence (i.e. the counterexamples), though. In other words, mathematical knowledge could be still deemed a priori, because it doesn't rely substantially on empirical evidence, but it is revisable, as Lakatos' work shows. On the other hand, we could think that Lakatos' views have an import on the epistemic status of logic too, as mathematics and logic share a common use of abstract concepts and categories, they both seemingly rely on rational processes having little to do with experience, and they seem to have special connection with one another. Therefore, if we accept that mathematics is rationally revisable for the reasons highlighted by Lakatos, we could expect the same line of reasoning to apply to logic in order to argue that logic can be rationally revised, as well.

The role of a posteriori evidence in Lakatos' view of mathematics and its application to logic doesn't necessarily challenge the apriority of mathematics or logic, as long as rational revisability is considered inoffensive for apriority. The discussion in Lakatos (1976) highlights how a priori knowledge is sensitive to a posteriori evidence, but it remains to be explained in what sense a posteriori evidence is a threat for the apriority of knowledge.

In the introduction, we have seen that a proposition is considered a priori iff it can be known or justified independently of experience and, to be clear about what counts as a priori knowledge, it is crucial that we are clear about what kind of a posteriori evidence is decisive for knowledge to be classified as empirical. One of the points made by those who have discussed the a priori is that certain kinds of experience are not relevant for invalidating the a priori status of mathematical or logical propositions. Checking a proof requires some amount of experience, i.e. someone (yourself or someone else) has to go through your proof and check whether each step of the derivation follows from the others.

Therefore, to some extent, revising a proof involves a posteriori activity, but it might be replied that here lies a confusion among different senses of a priori and a posteriori. It is true that checking a proof is a process taking place *after* the proof has been laid out (in this sense it *is* a posteriori) and that checking a proof requires that someone has the competence and the experience to draw proofs. However, there is no straightforward way to show that the process of checking a proof is not itself a priori, as it is a mental activity that doesn't rely on empirical evidence in any significant way. Consequently, it is not straightforward that checking a proof is enough to turn the proof into an a posteriori piece of knowledge.

2.2.2 Against the a priori. Williamson on empirical evidence

Cases like the proof checking fall under the category of rational revisability, which is usually held as being compatible with the a priori, and it is contentious whether rational revisability is really an interesting issue for the debate about the a priori.

A radical empiricist could follow Quine (1951) and reject the Fregean distinction between matters of meaning and matters of fact: as long as experience plays *any* role in the acquisition of knowledge or in the revision of knowledge, it makes that piece of knowledge empirical. However, to the radical empiricist it could be replied that the issue is much more nuanced than it seems and that whether a proposition is known a posteriori depends on the *role* that experience plays in acquiring the knowledge of said proposition. The traditional response to empiricism has been summarized by Williamson (2013), who follows Burge's[1] distinction between an *evidential* role and an *enabling* role of experience in acquiring and using concepts. (Williamson (2013): 292-294)

Williamson explains that experience is regarded as having an evidential role whenever it plays a direct role in acquiring knowledge. So, for example, to know that water freezes at 0°C, we need to perceive that the water is turning into ice and we need to perceive that the temperature has hit 0°C, e.g. by looking at a thermometer.

On the other hand, experience is held to have an enabling role whenever it makes us able to use correctly a certain word or concept. For instance, to infer from 'If the water has frozen, then the temperature must be lower than 0°C', and from 'The water has frozen' that the temperature is 0°C or less, doesn't require that we read the temperature on a thermometer. We simply apply an instance of Modus Ponens to the knowledge we already have, and infer that the temperature must be lower than 0°C.

[1]See Burge (1993).

Therefore, according to Williamson, to show that a proposition (suppose a mathematical or logical proposition) is known a priori, it is crucial to show that (i) experience has a merely enabling role in the acquisition of the knowledge of that proposition, and that (ii) experience plays no role in revising the knowledge of that proposition. If the two conditions are met for some propositions, we can claim to have a priori knowledge, whereas if we are not able to identify any proposition meeting the conditions, we should reject that there is anything like a priori knowledge.

Before we discuss the empiricist attack to the a priori, we might want to mention that Williamson criticizes the a priori/a posteriori distinction on non-empiricist grounds, taking into account not only the experience connected to the senses (*outer* experience), but also the experience obtained through introspection or reflection (*inner* experience). The consequences of taking inner experience to be on a par with outer experience will be addressed in more detail the last section. For now, I will take a traditional stance on experience, holding that it means 'sensory experience'.

It is traditionally accepted that (sensory) experience has a purely enabling role in obtaining a priori knowledge and, if it is to maintain its epistemic status, a piece of a priori knowledge shouldn't be revisable on empirical grounds. The topic of empirical unrevisability has been addressed by Field (2005), who distinguishes two definitions of the a priori called, respectively, *weak* and *strong* a priori, and relates them to the role that experience plays in knowledge acquisition. Weak and strong a priori will be the object of section 2.4.

Prior to addressing the two senses of the a priori discussed by Field, let's have a look at the empiricist view about logic.

2.3. Empiricism about logic

2.3.1 Putnam on quantum logic

Quine famously argued that in principle all knowledge is empirically revisable, but it wasn't until Putnam's famous paper in 1968 (reprinted in Putnam (1979), to which I will refer here) that the question of the empirical revisability of logic received much attention. Putnam's main question in his 'Is logic empirical?' is: is logic revisable in light of new empirical data? Putnam anticipates that the answer is 'yes', making of logic an empirical science as much as physics is. Putnam's article includes a discussion of the contribution of empirical science to the evolution of geometry, and its analogy with the case of logic.

First, Putnam draws a parallel with non-Euclidean geometry (Putnam (1979): 174-176), where the example is taken to show that a *prima facie* a priori science (geometry) has in fact turned out to be a posteriori regards the case of parallel lines. Contrary to the classical definitions of parallel lines, the General Theory of Relativity has shown that two parallel straight lines could well converge in the end and, according to Putnam, the idea that two parallel lines can converge challenges the supposed necessity of "a priori" propositions. In other words, as it is thought to be a necessary truth that parallel lines remain equidistant, because they have been defined as such, Putnam argues that the case of parallel lines shows that empirical discoveries can disconfirm propositions that were thought of as necessarily true (Putnam (1979): 176).

Another important example is provided by quantum mechanics, which has inspired quantum logic. Specifically, quantum logic doesn't validate $P \wedge (Q \vee R) \leftrightarrow (P \wedge Q) \vee (P \wedge R)$, nor $P \vee (Q \wedge R) \leftrightarrow (P \vee Q) \wedge (P \vee R)$ (distributive laws),[2] which makes it not only a non-classical logic, but an *empirically grounded* revision of classical logic (Putnam (1979): 184).

In light of the case of parallel lines, Putnam argues, if geometry can be proved to be empirically grounded, why should we resist the hypothesis that logic can be revised on empirical grounds as well? Related to the question of the empirical foundations of logic, there is the question about the factuality of logic: if logic can be revised on empirical grounds, would that mean that logical laws somehow capture the structure of reality (or can be expected to do so)? In other words, are logical theories different in any significant respect from, say, physical theories? Putnam answers that logic is as empirical as geometry and it is 'in a certain sense, a natural science' (Putnam (1979): 174).

Another point addressed by Putnam is the issue of conventionalism. Does the change of logic amount to a mere change in the conventions of language, as conventionalism seems to imply? The first brief reply to the question is that, if logical laws were just linguistic stipulations, there wouldn't be any way to know whether they are fit for rational inquiry (Putnam (1979): 188). The discoveries of quantum mechanics

[2] More in detail, in quantum logic the first biconditional fails left to right, and the second biconditional fails right to left.

highlight that there is a gap between the implications of the empirical results for the distributivity laws and classical logic. The problem can be solved either by presupposing that there are some problems at the methodological or theoretical level with quantum mechanics, or by concluding that quantum mechanics does disconfirm the classical laws of distributivity.

As apriorism can be formulated in terms of the view that logical laws are true in virtue of language alone, we shall analyze the question in more detail. One important issue raised by inferentialism about logic is the following: we might see the meaning of the connectives as imposed by the corresponding introduction- and elimination-rules (or introduction-rules alone). So, for example, the meaning of $\wedge$ would be specified by $P, Q \vdash P \wedge Q$, and $P \wedge Q \vdash P$, $P \wedge Q \vdash Q$. According to Putnam's take on the inferentialist view, a language without a connective that follows the inferential rules (classically) specified for it, doesn't contain the concept associated with said connective at all. To this, Putnam replies sceptically about the philosophical relevance of inferentialism:

> To mean 'or' e.g. a connective must satisfy such principles as: '*p* implies *p or q*' and '*q* implies *p or q*', simply because these formulate the properties that we count as 'the meaning' of 'or'.
> Even if this be true, little of interest to the philosophy of logic follows, however.
> [...] it does not follow either that a language which is adequate for the purpose of formulating true and significant statements about physical reality *must* contain a word *V* which obeys such-and-such patterns of inference, or that it *should* contain a word *V* which obeys such-and-such patterns of inference. (Putnam (1979): 188-189)

To the change-of-meaning argument, Putnam replies that the introduction- and elimination-rules for $\wedge$ and $\vee$ are the same in quantum and in classical logic; therefore, even from an inferentialist point of view, there is no change of meaning involved with respect to the connectives, but only a change in view with regard to the distributivity laws.

To save classical logic from revision, one might be tempted to dismiss the empirical evidence with all the sorts of expedients, such as pointing out that there are important methodological mistakes in quantum mechanics, like undetected variables, mistaken reports, gaps in the observations registered, etc. Alternatively, one might reject quantum physics as a truthful scientific theory and so deny that we should revise classical logic in light of quantum mechanics because it is plainly wrong. However, Putnam argues, instead of resisting empirical evidence, it seems more reasonable just to bite the bullet and admit that classical logic is wrong because it doesn't reflect reality.

2.3.2 Replies to Putnam

Unsurprisingly, Putnam's empiricism about logic has been challenged in many ways. For example, Dummett, in his paper also called 'Is logic empirical?', asks whether the empirically grounded logic should really replace classical logic, and answers that logical revision on an empirical basis changes the *meanings* of the connectives, similarly to logical revision on non-empirical grounds.

Dummett argues that quantum logic, as presented by Putnam, in fact introduces two new connectives $\wedge$ and $\vee$, and the analogy with the straight lines confirms that the revision boils down to a shift of meaning, contrary to what Putnam tries to prove. In the example of the straight lines, Dummett argues that Putnam

> [...] insist[s] on two points: (i) that there is enough in common between the uses of 'straight line' in the two contexts for the use of the same expression not to be a mere equivocation; and (ii) that we have not merely shifted the label 'straight line' from one set of paths to another. [...] Point (ii) is to the effect that this is *not* what happens in the geometrical case: even if 'straight line' has changed its meaning, we cannot, having adopted Riemannian geometry, use *any* expression with just that meaning which 'straight line' used to have when we adhered to Euclidean geometry. (Dummett (1976): 284)

In fact, according to Putnam, when a meaning shift is just a re-labelling, we are still able to express the old meaning, but in the case of geometry, we aren't. On the other hand, Dummett argues that the meanings of $\wedge$ and $\vee$ in quantum logic are held alongside their classical meanings by Putnam, as Putnam himself cannot help but use classical logic.

Dummett summarizes Putnam's argument roughly as follows: suppose A stands for the position of a particle at a given time, and that B_1, ..., B_n stand for all the possible momenta the particle has. As we cannot know both the position of a particle and its momentum, the statement $A \wedge B_i$, for each i among B_1, ..., B_n is false. But as we know that A is true (suppose we measured the particle's position and we believe the measurement to be accurate), classical logic would take us to claim that $B_1 \vee ... \vee B_n$ is false, i.e. that the particle has no momentum. However, as we know, the particle has momentum, i.e. $B_1 \vee ... \vee B_n$ is true, and in fact also $A \wedge (B_1 \vee ... \vee B_n)$ is true, but from this we cannot derive $(A \wedge B_1) \vee ... \vee (A \wedge B_n)$, as we know that $A \wedge B_i$ is false. (Dummett (1976): 271-273)

However, Dummett argues, if Putnam wants to be a realist about quantum mechanics, he is forced to keep the laws of distributivity; conversely, if Putnam wants to reject bivalence and distributivity, he cannot be a realist about quantum mechanics. Why? Well, Putnam believes that any particle has both momentum and a position at any given time, but he claims that we just cannot know them both, otherwise we would know a logical contradiction. But the contradiction is a contradiction *only* in a factual (quantum-mechanical) sense, and

> the notion of truth to which Putnam appeals when he defends a realist view of the quantum-mechanical system, and when he says that there is a true statement about the momentum which, knowing a true statement about the position, we do not know and cannot know, is a notion which obeys classical, two-valued principles. (Dummett (1976): 273)

In fact, Dummett argues, Putnam's argument relies on the fact that if $B_1 \vee ... \vee B_n$ is true, one of the disjuncts must be true, therefore Putnam relies on the classical meanings of $\vee$ and $\wedge$ for his argument to follow and in turn, he relies on the distributivity laws.

Therefore, the analogy between the case of geometry and the case of logic is undermined, and the logical revolution of quantum logic is ultimately a re-labeling of the meanings of $\wedge$ and $\vee$. Dummett's response to Putnam invokes the principle that any attempt to revise logic is an attempt to change the meanings of the logical constants, ruling out any possible logical revision as a case of semantic disagreement.

Dummett's reply to Putnam seems to commit one to the view that logical revision is more than a disagreement about meanings, which seems a rather hard bite to swallow. Therefore, if we take logical revision to be an example of genuine theoretical disagreement, we should look elsewhere for replies to empiricism.

Another typical response to Putnam's argument involves the role of empirical methodology in gaining experience, and it will be discussed in the next section. For now, it will suffice to say that the methodological argument roughly goes as follows. It could be argued that some a priori truths are crucial for collecting data and making sense of the empirical evidence collected. Therefore, before we are actually called to revise our a priori methodological principles in the face of new experience, we must bear in mind that the experience has been acquired through those very principles that it seems to defeat.

On the other hand, Field argues that being aprioristic about logic doesn't commit one to holding that logic is unrevisable come what may (he calls this *non-dogmatic apriorism*). We will see in section 2.6 how contemporary empiricism replies to the methodological objection against empiricism, whereas Field's apriorism will be the object of the next section.

2.4. Negative characterizations of the a priori

Before we discuss weak and strong notions of the a priori, it should be made clear that there are two main characterizations of the a priori. There are positive definitions of the a priori, i.e. definitions that specify under what conditions a proposition, p, is known a priori, and under what conditions one's entitlement to believe such a proposition is a priori; and negative characterizations of the a priori, i.e. definitions that specify what needs to be absent from the warrant of p, for p to be known a priori.

We have already mentioned that the common definition of the a priori states that for a proposition to be known a priori, it has to be known independently of experience. 'Independently of experience' characterizes this definition as a negative one, as it specifies that p is known a priori if and only if experience doesn't play a role in our entitlement to know that p.

Starting from the negative definition of the a priori, Field (2005) distinguishes two senses of the a priori: propositions that one has an priori entitlement to believe in a *weak* sense, and propositions that one has an a priori entitlement to believe in a *strong* sense.

Assuming that 'someone knows that p only if p is true, she believes it, and she is *entitled* to believe it; the issue of a priority concerns the kind of entitlement that is in question'. (Field (2005): 70) For a proposition, p, to be a priori in the weak sense, one needs to have *gained* entitlement to believe p independently of experience. In other words, there is no experience that has made one entitled to believe p.

On the other hand, for a proposition, p, to be strongly a priori, p must be *unrevisable* on empirical grounds. In other words, no experience counts as defeating the entitlement to believe p, although this doesn't mean that, to be a priori, a proposition has to be unrevisable *tout court*. If p is revised by reason alone, it might well still be a priori, whereas the kind of revisability which could really cast doubts on the apriority of a piece of knowledge is empirical.

Field considers the strong sense of the a priori as philosophically more interesting, and he analyzes two versions of it. According to Field, one can hold that some knowledge p cannot be revised in two important senses. In the first sense, p couldn't possibly be revised under any circumstances, i.e. it is *impossible* that p can be revised. In the second sense, there is a remote possibility that p can be revised, but there is no actual evidence for that possibility.

Field argues that to claim that logic is unrevisable in an epistemic sense, i.e. that logic cannot possibly be revised on empirical grounds, is far-fetched and difficult to prove, so a less dogmatic view of the revisability of logic is more plausible. Instead, there is a remote possiblity that logic can be revised empirically, e.g. in light of quantum mechanics.

In addition, Field argues that we have a default entitlement to some beliefs, i.e. an entitlement to regard some propositions as true without having to put them to the test. In Field's words:

> Probably the best view is that we simply have an attitude of regarding some beliefs as entitled under some circumstances, others not; and we regard some of them as entitled in absence of evidence for or against, even though there might someday be evidence that disconfirms them. (Field (2005): 75)

Could it be that we believe logical and mathematical propositions by default?

We might think that any such default entitlements require a source. For example, we might think that the source of the entitlement could be found in the meanings of the logically-relevant concepts, but Field considers a different option. The entitlement to the logical truths could, according to Field, depend on the possibility of genuine disagreement.

To introduce the point, Field distinguishes the case of disagreement about a methodology from the case of the disagreement about mathematics. In the former case one can value, say, inductive methodologies over counterinductive ones. There is a genuine disagreement about what methodology is the most reasonable one, the most apt for constructing reliable theories of the world, and so on. In the case of empirical methodology, we might think that there must be a way to justify the choice of a methodology over the others, i.e. a source of our entitlement to a methodology.

The case of disagreement in mathematics is very different, Field argues. Mathematics is the sort of science where a genuine disagreement cannot exist and, in fact, when two consistent theories seem to rival each other, they are just about different subjects.

The case of logic, Field concludes, is more akin to the case of empirical methodology than to the case of mathematics. In fact, we seem to need a justification to regard as correct some basic beliefs about logic. Especially in the case of conflicting logics, we think that there is something making it the case that a logic is correct and the other one is wrong.[3] Although it is difficult to identify what it is exactly that rival logics disagree on, Field suggests that rival logics disagree about the rules of rational belief (as it will be discussed later).

If the matter of the disagreement is really the principles of rational belief, what would entitle us to use the rules of a certain logic instead of the rules of another logic? Some think that meaning, inferentially or truth-theoretically intended, is the source of this entitlement. However, justifying our entitlement through the truth-theoretic meaning of the logical expressions would be circular. According to Field, the case is different with inferential semantics, i.e. with a semantics that defines the connectives by the corresponding inferential rules. As on an inferentialist account of logic, accepting some basic logical beliefs about inferential rules is strictly connected to accepting the meaning of the corresponding connectives, this connection provides a justification for the corresponding principles, Field argues.

[3] Supposing that we are not pluralist about logic, of course.

If we disallow circular justification, we are urged to look for a source of entitlement elsewhere. However, there is no proper source of entitlement for a logic, according to Field, because we cannot expect logic to be justified through arguments, nor can we expect it to be justitifed through non-argumentative forms of justification either. What in fact we have is a default entitlement to our beliefs that a certain logic is correct, which might nonetheless be called into question at times. For instance, our entitlement to a logic is called into question when our favourite logic is challenged: there we need to show why, say, LEM is a valid principle in the face of phenomena that seem to disconfirm its validity, such as vagueness.

Field's discussion of the entitlement to basic logical beliefs is reminiscent of Boghossian's views on the entitlement to use logical rules (which is discussed in the next section) and the debate about epistemic entitlements, which has been the object of extensive investigation (see, e.g. Dretske (2000) and Wright (2004) for a thorough discussion on this). Roughly speaking, epistemic entitlement is the kind of non-inferential and apparently unsupported warrant that one has toward certain propositional attitudes. The beliefs held through epistemic entitlement can come from, e.g. perception, memory, introspection. Among the beliefs we are epistemically entitled to, there are also some propositions we endorse a priori, e.g. logical truths. This raises the issue of whether a positive characterization of the a priori is possible. In other words, can we define apriority in terms of its connection to some rational faculty or to some other philosophically relevant concept?

2.5. Positive characterizations of the a priori

The a priori has been positively characterized mainly in two ways: (i) as knowledge gained through rational intuition (see Bealer (1998); BonJour (1998)), and (ii) as knowledge that is true and justified in virtue of language alone.

Intuition in general is described by BonJour (1998) and Bealer (1998) as the conscious experience such that if you have an intuition that p, it seems to you that p. To this extent, intuitions are explained as propositional attitudes which are nonetheless distinct from beliefs and, at the same time, also different from perceptions. Intuitions are different from perceptions mainly because the seemings involved in intuitions are intellectual seemings, rather than sensory ones. In turn, philosophically relevant intuitions need to be distinguished from other sorts of intuitions, as the intuitions that are relevant to philosophical inquiry are a priori intuitions, which manifest themselves as necessary (Bealer (1998)). Similarly, BonJour (1998) describes rational intuition as a propositional attitude such that if you have the rational intuition that p, it seems to you that necessarily p.

Unfortunately, the account of rational intuition provided by Bealer and BonJour faces an immediate problem: if a rational intuition that p is an a priori intuition, i.e. a seeming that necessarily p, it is not clear how the rational intuition account could give a definition of the a priori avoiding circularity.

The other positive characterization of the a priori defines it in terms of analyticity, as Boghossian (2000) does.

Similarly to the principles of mathematics, logical rules are deemed as justified/known a priori. According to the traditional negative characterization of the a priori, if knowledge of logic is a priori, it cannot be defeasible nor revisable on empirical grounds.

Boghossian asks two separate questions about how it is possible that we have logical knowledge. On one hand, there is the question about what justifies the belief that a certain inferential rule, say Modus Ponens, is valid. On the other hand, Boghossian asks what makes us entitled to use a logical rule in our derivations or reasoning. It is worth pointing out that Boghossian sticks to a specific use of 'justification' and 'entitlment', where 'justification' applies to beliefs, and 'entitlement' applies to the application of rules.

First of all, Boghossian raises the question of how we could justify our beliefs in logical rules. One can think that the belief that, in Boghossian's example, MP is valid could be justified somehow and ask whether the justification could be inferential or not. If MP cannot be justified inferentially, i.e. resorting to some other sort of rule that proves that MP is valid, then we need to show that there is some special intuition, e.g. rational intuition, pointing to the validity of MP. Rational intuition represents the rock bottom of the justification process, a sort of rational enlightenment, so one cannot just ask that rational intuition be justified itself by something else. There is another option, though. One might think that logical rules are justified because one cannot see any counterexamples to them. However, Boghossian quickly dismisses the option, arguing that not being able to recall any counterexamples to a rule is just a way to camouflage an inferential account of the justification of logical rules. Perhaps it is less than straightforward that looking for counterexamples bears epistemically on rule inferentialism, but I won't address this question here.

The issue of the circular justification of logic is discussed at length by Boghossian (2000) who, after having rejected the option that the justification of a logical rule can be non-inferential, discusses the inferential case. The first question coming to mind is therefore: can the justification of the validity of a logical rule be non-circular? Clearly not. Suppose you want to justify the validity of a derived rule, such as Disjunctive Syllogism. In that case, you will start a justification regress to more primitive rules, such as ∨-E, until you reach the point where you use the very same primitive rules to justify themselves.

On the proof-theory side, the issue of circular rule justification has been addressed by the question of justification in inferentialistic terms, as in Read (forthcoming). Read argues that Introduction-rules determine the meaning of the corresponding connectives and are therefore self-justified, whereas the Elimination-rules are justified in virtue of the corresponding I-rules and of the principle of General Harmony, which prevents the connectives from falling into *tonk*-like problems.[4] So, to take up Boghossian's example about MP, according to Read (forthcoming), MP being $\to E$, it is justified by Conditional Proof, i.e. $\to I$, which is self-justifying.

As circular justification is traditionally not welcome, Boghossian in his paper asks what is so bad about it. Is it really unadmissible that logic or mathematics are based upon some sort of circular justiifcation? Most of the effort in Boghossian's paper is devoted to defeat the prejudice against circular justification and to prove that in certain contexts, namely, logic, circular justification is the best we can get. To Boghossian, giving up rule-circular justification would make any a priori justification of logical rules impossible and it would pave the way to skepticism. Therefore, Boghossian concludes that we should ultimately bite the bullet and accept that the foundations of our logical knowledge rely on self-justification.

The empirical justification of logic would still be an option, but Boghossian argues that there are no clear reasons to prefer an empirical justification over an a priori justification of the inferential rules. In fact, Boghossian argues, suppose we are in search of an empirical justification of logical rules, and suppose that we want to compare how different theories of the world and the logics underlying them fare in the face of empirical evidence. Boghossian observes that we would need a meta-logic to compare the logics underlying the theories, pushing the circular justification from the object-level to the meta-level. Therefore, Boghossian concludes, if rule-circular

[4] I direct the reader to the first chapter for further discussion of Proof-theory and of the Principle of Harmony.

justification is a problem, opting for empiricism doesn't help with it and, as we cannot do without a justification for logical rules, circular justification is better than none.

On this account, logic is a priori in a weak sense, as experience doesn't play a role in the acquisition of logical knowledge or, to put it in Williamson's terms, experience has a merely enabling role. However, whether Boghossian would endorse a strong view on the apriority of logic or not remains an open question. That the rules of logic are only justifiable in a circular fashion certainly means that experience is not involved in any substantive way in the justification of a rule R, but it doesn't show that logical rules cannot be defeated on empirical grounds. A rule R could ultimately rely on itself for its own justification, but it is still conceivable that we revise our belief about R on the basis of the discoveries of, say, quantum mechanics. Therefore, R wouldn't be a priori in a strong sense, but it would be more akin to a default-entitled rule *à la* Field, in the sense that we assume that we are entitled to know and use R, but it is conceivable that R be defeated on the basis of some relevant experience.

In the previous two sections, we have examined some among the most important views on the a priori, dividing them into two main groups: those providing a negative definition and those providing a positive definition of the a priori. We have seen that in both cases, there are convincing reasons to think that the only justification that we can obtain of a priori rules, such as the inferential rules, is circular, so we need to *assume* that we are entitled to use those rules in our reasoning at least most of the time. On the other hand, being entitled by default to apply the rules doesn't rule out that the rules are (virtually, at least) fallible and revisable.

In the next section, fallibilism about the a priori will be addressed through Bueno's criticism of Field's aprioristic fallibilism.

2.6. Contemporary empiricism

Against Field, Bueno argues that Field's fallibilism is ultimately incompatible with apriorism about logic.

Bueno, in opposition to Field's fallibilist apriorism, presents a view dubbed *fallibilist empiricism*. Fallibilist empiricism displays no substantial difference with what we could call apriorism in a weak sense, following Field's definitions of the a priori.

As we have seen, that some knowledge is a priori in a weak sense means that one is entitled to that knowledge independently of experience. I call the corresponding view about logic *apriorism in a weak sense*. Similarly, fallibilist empiricism holds that 'although the legitimacy of logical rules does not depend upon empirical evidence, logic is empirically defeasible' (Bueno (2010): 108).

Bueno argues that there is no clear reason for preferring fallibilist apriorism over fallibilist empiricism. He addresses three arguments presented by Field in his defense of apriorism and counters each of them, trying to show that the implausibility of Field's arguments makes the fallibilist view incompatible with apriorism. The arguments in favour of the apriority of logic presented by Field revolve around, respectively, (i) the argument that empirical evidence is ultimately irrelevant to logical revision; (ii) the argument that logic must be a priori if it is non context-dependent; (iii) the argument that logic and geometry cannot be treated as if they were parallel cases of empirically revisable subjects. The third argument is specifically against the famous argument in Putnam (1979) that we have already discussed, and we will address it last.

I will start following the order of the arguments and their responses in Bueno's paper.

In Field (1998): 3, Field addresses the question 'What does it mean to revise a logic on empirical grounds?' or, more specifically, 'What does it mean to revise logical principles on empirical grounds?'. A first shot could be that revising a logic empirically is what happens when we discover that Santa doesn't exist and consequently decide to revise our logic to include non-referring terms. However, Field replies, this is not a genuine case of revision on empirical grounds, as the discovery made through empirical evidence that there are non-referring terms 'is simply serving to bring our attention to a conceptual possibility that we should have already recognized.' (Field (1998): 3) Field dismisses the example as uninteresting, because empirical evidence has a marginal role in demonstrating that we need to accommodate non-referring terms into our logic. Therefore, we need some sort of discrimination between cases like the example of the non-referring terms and cases like quantum logic, where empirical evidence is appealed to more seriously.

To this argument, Bueno replies as follows: if the empirical evidence that Santa doesn't exist is dispensable for the revision of logic because it just points out some conceptual possibility we have failed to recognize, the same should apply to quantum mechanics. The evidence from physics experiments against the distributivity laws could be seen as irrelevant, as it points out to some conceptual possibility we should have already recognized. However, Bueno argues, this is plainly implausible, and the argument against the relevance of empirical evidence to logical revision can be pushed further. Suppose we, like Field, claim that an empirical phenomenon brings to our attention a possibility that we should have recognized independently of experience. Then, Bueno argues, we would also be committed to the claim that we can derive all the scientific discoveries a priori. Therefore, Field's argument ends up proving that scientific theories are empirically indefeasible too. On the contrary, Bueno argues that the revisions to our theories flowing out of the scientific discoveries are not conceptual at all. Indeed, the revisions are empirical as much as the phenomena that motivated them.

However, Bueno's reply to Field's first argument seems to be very much in the spirit of Putnam (1979) and, as it aims to undermine the apriorist's argument on the basis of radical empiricist assumptions, it is very theory-laden.

The second argument Field outlines to show that logic is a priori is centered on the universality of logic: we need logic as a methodology for empirical research. Therefore, logic cannot be context-dependent and it must be a priori. In brief, the argument to the universality of logic goes as follows: empirical methodology needs an underlying logic which enables us to collect and make sense of the empirical evidence, so there must be some correct logic that meets our epistemic and methodological needs.

Bueno responds that there is no clear reason why logic should be deeemed as a priori in virtue of being what underlies empirical theories. There might be a manifold of good logics, each adequate to a certain context or domain.

For example, classical logic might be adequate in certain contexts but not in others, where it would be appropriate to use intuitionistic or quantum logic instead. In certain contexts it might be appropriate to apply logics that are clearly empirically revisable, such as quantum logic, whereas in other domains it may be better to apply logics that are not considered empirically revisable by many, such as classical logic.

Bueno argues that the logical landscape emerging from this pluralism of applications and domain-dependency doesn't force the conclusion that classical logic has a status different from the status of other logics in virtue of not being (allegedly) empirically revisable. The logic underlying the methodology can change as much as the theory can.

However, the empiricist is not committed to the view that each of the logics which are part of the plurality is empirically defeasible. Some of them might not be empirically revisable but only rationally revisable, and this wouldn't undermine the empiricist stance anyway, because empiricism about logic holds that *some* logics are revisable on empirical grounds, not that *all* of them are.

Bueno's reply to the universality argument seems to address the question 'Should we apply classical logic to all the domains?', rather than the question 'Should we use *one* logic for our inquiry methods in all domains?'. Suppose we are interested in the study of certain nanoparticles that sometimes display a certain property, F, and suppose that, upon accurate investigation, we find out that some of the nanoparticles have F, some don't, and for the remaining nanoparticles it is uncertain whether or not they have F. Therefore, we conclude that, given the state of our technology and research, a logic including LEM among its laws can't be used to describe the behaviour of the nanoparticles adequately.

However, when we consider our study of the nanoparticles and we plan the experiments we are going to conduce to investigate the properties of the nanoparticles, we *assume* that either the nanoparticles have the properties F, G, H, ..., or they lack them. In other words, we do apply LEM in our investigation method, tacitly giving ourself a default entitlement to apply LEM to collect the experimental data. The logic underlying our methodology could nonetheless be open to revision, although revising our methodology might take much longer than revising the logic we apply to a certain domain.

Finally, about the third argument, Field tries to show that the parallel between logic and geometry doesn't hold. The fundamental difference between logic and geometry, according to Field, is that whereas logic underlies our empirical theories, geometry doesn't. Therefore logic, unlike geometry, has a central role with regard to our knowledge. For this reason, it makes sense to argue that geometry is empirically revisable, but it is unadmissible that logic is genuinely revisable on empirical grounds.

However, Field leaves open the possibility that logic could be empirically revisable in an epistemic sense, i.e. that it is conceivable that logic be revised on empirical grounds but that there is no actual evidence that might suggest a logical revision.

Bueno replies that Field's third argument overlooks a distinction common to geometry and logic, i.e. the distinction between pure logic/pure geometry and applied logic/applied geometry. Pure logic is the abstract study of consequence relations in general and it can be applied to derive consequences about objects in particular contexts, which is where applied logic arises. It is at the level of applied logic that we might want to revise the logical principles, Bueno argues, as we could recognize that the empirical evidence calls for a revision of the general principles.

In the reply to the third argument, the clash between two opposite assumptions (the apriorist's and the empiricist's) shows up again, making it seem as if Bueno and Field are talking past each other. However, it is doubtful that Bueno really proves that empirical fallibilism is to be preferred over aprioristic fallibilism, partly because the two fallibilistic views are similar in many respects, and partly because Bueno's second reply seems to rely on a change of subject. For these reasons, the contemporary empiricism represented by Bueno doesn't seem to be a big threat for apriorism. In the next section, a more pernicious view is discussed, that tries to dissolve the distinction between a priori and a posteriori knowledge.

2.7. Rejection of the a priori/a posteriori distinction

The difference between contemporary empiricism, as presented by Bueno, and the fallibilist apriorism held by Field is very thin, indeed almost unrecognizable. So, it makes sense to ask whether the distinction between a priori and a posteriori is really philosophically significant. Similar worries have been raised by Williamson (2013), who denies that the distinction between a priori and a posteriori is really central to the philosophical discussion. As we have seen in the first section, Williamson claims that traditional discussion on the a priori has tacitly assumed that there is a substantial difference between an evidential role of experience and an enabling role of experience.

Nonetheless, Williamson argues that the distinction between the two roles of experience is much more blurred than tradition has taken it to be, as there are no clear-cut cases in which experience plays a merely enabling role in the acquisition of concepts. Therefore, Williamson concludes, the distinction between a priori and a posteriori is not substantial.

To support his views, Williamson suggests that we take two cases that seem clear cases of a priori and a posteriori knowledge, respectively, namely the sentence 'All crimson things are red' and 'All recent volumes of *Who's Who* are red'. But now suppose that someone, a boy called Norman in Williamson's example, learns the meanings of 'red' and 'crimson' by ostension and doesn't know that they are correlated. When asked whether 'All crimson things are red', Norman imagines a sample of crimson and a sample of red, and assents to the statement. Norman is then asked whether 'All recent volumes of *Who's Who* are red' and having seen the volumes, having acquired competence with the words 'volume' and 'recent', after reflecting for a moment, Norman assents to the statement about the recent volumes of *Who's Who*.

Williamson argues that the mental activity underlying Norman's two responses to the statements are essentially the same, as in both cases Norman imagines samples of the objects included in the sentences. Therefore, Williamson asks, if the cognitive processes involved in acquiring a clear case of a priori knowledge and a clear case of a posteriori knowledge are the same, how can the distinction between a priori and a posteriori be really meaningful? (Williamson (2013): 296)

Another point undermining the a priori/a posteriori distinction lies in how we define experience. In the second section, we had seen that Williamson puts inner experience on a par with outer experience. Williamson argues that including inner experience into the use of 'experience' has some worrying consequences for the supporter of the a priori/a posteriori distinction. On one hand, one might worry that the inner experience needed to perform some mathematical calculation in one's head could turn out to have an evidential role, making the mathematical calculation a piece of a posteriori knowledge. But more importantly, the outer experience could turn out to play an enabling role in empirical science. As Williamson puts it,

> [...] part of the evidence that a massive comet or asteroid collided with the Earth about 250 million years ago is said to be that certain sediment samples from China and Japan contain certain clusters of carbon atoms. That those samples contained those clusters of atoms is a nonpsychological fact. Of course, in some sense scientists' outer experience played a role in their access to the fact. But, by analogy with the logical and mathematical cases, the relevant evidence is not the psychological process of undergoing those outer experiences, but rather the non-psychological physical facts to which that process enables us to have access. The role of the outer experience is purely enabling, not evidential. If so, what would usually be regarded as paradigm cases of a posteriori knowledge risk reclassification as a priori. (Williamson (2013): 294)

The issue raised by Williamson is a serious one, and it shouldn't be taken too lightly. The distinction between a priori and a posteriori knowledge relies on the definition of 'experience' we give, but Williamson asks why inner experience should not count as experience in the relevant sense. On the other hand, if we did, the distinction between what is known a priori and what is known a posteriori would become more confused than it traditionally has been taken to be.

More importantly, Williamson's argument aims to show that even clear-cut cases of a priori and a posteriori knowledge have the same underlying mental process, legitimating the question 'If it is not the epistemic process underlying the two sorts of knowledge that makes the difference between a priori and a posteriori, what does the job?'. This is perhaps the main threat to the a priori. Nonetheless, Williamson suggests that philosophy can dispense from the traditional epistemological distinction without much damage.

Williamson's argument relies heavily on the radical assumption that inner experience should be treated on a par with sensory experience. Without assuming that introspection, reflection, memory, etc. must be treated as empirical sources of evidence on a par with perception, Norman's knowledge of the two propositions would turn out to be of very different kinds. In fact, to know that 'All the recent volumes of *Who's Who* are red' Norman presumably relies on memory, imagination and the perceptual experience he had when he saw the volumes in question, where the visual experience plays an evidential role in Norman's assent to the proposition.

On the other hand, to know that 'All crimson things are red', Norman relies on memory, imagination and his rational abilities to grasp the concept 'red' and 'crimson'. The visual experience Norman had when he first saw a sample of red and a sample of crimson *enabled* him to form the concepts 'red' and 'crimson', but Norman understands that 'All crimson things are red' is necessarily true in virtue of *those* concepts, rather than in virtue of the original visual experience.

Therefore, it is hard to prove that there is no significant distinction between a priori and a posteriori knowledge unless we make the rather bold assumption that inner experience counts in the same way as sensory experience, epistemically speaking.

Mathematics and logic might not be seriously affected by a change in their epistemological status, as Bueno argument purports to show, and sometimes aprioristic and empirical views might be divided by a line so thin that makes the debate look almost flimsy, but we must concede that whenever we engage in reasoning, we *assume* that we have reason to apply certain inferential principles, which might well be a priori *and* open to empirically grounded revision, even if in remote circumstances. For these reasons, it seems plausible that logical norms be recognized as rules that we are entitled to apply by default, as suggested by Field (2005). In the last chapter, I will give more reasons in support of the default-entitlement view.

2.8. Conclusion

In this chapter, we have addressed the question 'Is logic a priori?', which is relevant to the question 'Is logic a source of a priori norms?', and we have surveyed the main views on the epistemic status of logic. First, we have considered the definition of a priori knowledge, and observed that it raises a number of questions; in particular, defining *a priori* knowledge as knowledge that is gained independently of experience pushes us to distinguish two roles that experience plays in acquiring knowledge, namely *enabling* and *evidential*.

Traditionally, apriorists have argued that whenever experience has a merely enabling role in producing some kind of knowledge, said knowledge is considered as a priori. On the contrary, experience plays an evidential role in the acquisition of a posteriori knowledge, as in the natural sciences.

On the other hand, the empiricist often denies that there is a distinction between the enabling and evidential roles of experience. However, an empiricist view more relevant to our discussion is the modern empiricism of Putnam, who has argued that since logic is open to revision on empirical grounds, it cannot be called *a priori*. As we have seen, Putnam argues that the finding of quantum mechanics have a direct impact on logic, showing that the distributive laws don't hold in the realm of quantum physics. Therefore, Putnam argues, as empirical discoveries shape our theory of the world, and as logic is aimed at truth as much as the other sciences, if an empirical finding provides evidence against a law of classical logic, that evidence provides grounds for rejecting the law in question on a logical-theoretical level.

Many have replied to Putnam's attack to the apriority of logic, with Dummett responding that the attempt to revise logic relies on a meaning change, and with Field arguing that a priori truths are essential in the methodology of research, as we collect empirical data and make sense of them relying on inference rules. Therefore, to revise a priori truths, such as logical rules, on empirical grounds we rely on empirical evidence that has been collected and interpreted on the basis of those very a priori truths.

We have also addressed Field's negative account of a priori. We are familiar with the traditional definition of a priori knowledge as knowledge gained independently of experience, which is a negative characterization of the a priori. Field takes the discussion a step further, clarifying the sense in which a priori knowledge should be independent of experience. As we have seen in section 2.4, Field distinguishes two senses of the a priori, i.e. *weak* and *strong*.

In a weak sense, a proposition is known a priori whenever the knowledge of p is gained independently of experience, as in the traditional definition. On the other hand, a proposition is strongly a priori whenever it is unrevisable on empirical grounds. In turn, Field distinguishes two "strictness" levels of the empirical unrevisability: either it could be impossible to empirically revise a proposition *tout court*, or it could be remotely possible to revise a proposition on empirical grounds, even though we currently don't have any actual evidence for that possibility. Field argues that proving that logical laws are unrevisable in the first sense (impossible to revise) is too difficult, and that is a reason why we should be more open to a non-dogmatic apriorism, i.e. an apriorism that doesn't rule out empirical revision across the board.

The debate on the epistemic status of logic has included also the issues of the justification of our beliefs in the logical laws and our entitlement to use them. Some, such as Field, have argued that our entitlement to apply logical rules has no proper source; in fact, we have what could be called a default entitlement to use the laws of logic for our reasoning purposes. A similar view of the entitlement to logic has been endorsed by Boghossian, who argues that we need to accept circular justification as the only justification logical truths can have; therefore, Boghossian argues, we have a sort of justification for our beliefs in logical truths by default.

To these modern aprioristic views on logic, Bueno opposes his fallibilist modern empiricism. Bueno's account is very similar in spirit to Field's account of logic as (remotely) open to empirical revision, except that Bueno argues that although logical truths are not empirically justified, they are empirically revisable. However, we have seen that Bueno doesn't provide conclusive arguments against Field's position, failing to prove fallibilist empiricism.

We addressed another threat to logical apriorism in the last section, where we have discussed Williamson's (2013) argument against the a priori/a posteriori distinction. There, Williamson argues that a priori and a posteriori are not significant philosophical categories, as they are much more blurred than the philosophical tradition has taken them to be. However, we have seen that Williamson's argument relies on a very substantial assumption, i.e. that inner experience (related to mental states, such as reflecting, remembering, etc.) should be treated on a par with sensory experience. Without this assumption, Williamson is unable to establish that there is not a significant difference between a priori and a posteriori knowledge and, as the assumption is quite controversial, I think it shouldn't be taken too lightly.

In conclusion, we have seen that there is no serious threat to the apriority of logic, at least not from the modern empiricist views we have discussed nor from the views that try to dissolve the a priori/a posteriori distinction. In light of this, I have argued that a fallibilist apriorism, which takes logical rules to be justified by default, is the most plausible option.

CHAPTER 3

Current accounts of normativity

Abstract. In this chapter, I want to address current theories of the normativity of logic. First I will talk about how the scope of logic is taken to be constrained by necessity and formality, the characterization of which is far from being uncontroversial. MacFarlane addresses the issue of the formality of logic in his (unpublished, but available on his webpage) doctoral dissertation and the issue of the normativity of logic in a more recent paper, outlining an account of logic which stresses the importance of subjective acknowledgment of the inferential patterns in evaluating reasoning. I motivate the investigation of the normativity of logic by addressing a paper by Elqayam and Evans that discusses the normative approach taken by the psychological study of deductive reasoning to highlight how much of said study is jeopardised by implicit applications of *is-ought* inferences. I show how E&E's argument relies on a controversial claim that cannot be accepted in light of a close analysis. Then, I discuss the bridge principles introduced in MacFarlane (2004) and agree with MacFarlane that some of them (dubbed *Wr+, Wr-,* and *Wo-*) describe *prima facie* plausible ways to frame the constraints imposed by logic on informal reasoning. Then, I analyze the deontic operators used in the bridge principles, distinguish five senses of *ought* and make a hypothesis about the sense in which *ought* is employed in *Wo-*. Then, I suggest that one of the operators included in the bridge principles (i.e. *has reason to*) can be made sense of through the minimal deontic logic developed in Chellas (1974, 1980).

3.1. Introduction

In this chapter, I introduce the topic of the normativity of logic highlighting the connection between formality and normativity, then I make some hypotheses about the deontic operators in the normative principles.

In drawing the borders of logic, philosophers have to take into account many (competing) factors; they try to sail safely between the philosophical quest for formal rigour and the commonsensical quest for an intuitively appealing characterization of logic. The demand for an intuitively plausible explanation of mathematical structures is widespread in logic. Kripke's construction and ternary relations in relevant logic are examples of mathematical frames which have received philosophical and "intuition-friendly" interpretations, without which they would have remained significantly obscure to many philosophers.

As can be immediately noticed as soon as one looks at an encyclopedia entry or at a basic philosophy of logic textbook, the main characteristics of logic which are mentioned most often are *truth preservation* of inferences, *necessity* and *formality.* Leaving aside truth preservation, the interplay between necessity and formality allows at least five combinations, i.e.

1. Necessity constrains validity.
2. Formality constrains validity.
3. Formality and necessity have the same power of constraining validity.
4. Necessity is "reduced" to formality.
5. Formality is "reduced" to necessity.

Of the five combinations above, at least three have been discussed in the literature. The first option on the list corresponds to what Etchemendy calls *representational semantics*[1], where a proposition is valid iff it is true across all the possible worlds. This in turn amounts to constraining logical validity by necessity (truth across all possible worlds).

The fourth combination corresponds to Tarski's take on logical validity.[2] In fact, to characterize validity in mathematical terms, Tarski first established that the non-logical constants of the sentences be replaced by variables to obtain *sentential functions* (in Tarski's words) or *forms.* Then, he defined the notion of a model of a class of sentences K as an arbitrary sequence of objects which satisfies each sentential function of the class K, or (as modern logicians would put it) an interpretation which makes each sentence of K true. Finally, he defined validity employing the notion of *model.* In Tarski's words,

> We say that the sentence X follows logically from the sentences of the class K if and only if every model of the class K is at the same time a model of the sentence X. (Tarski (2002): 186)

[1] Presented in Etchemendy (1990).

[2] At least according to Etchemendy's construal of Tarski on logical validity.

Therefore, by defining validity in terms of substitution, Tarski set up a definition of validity in which necessity is reduced to formality. We will see how this category has a correlate in one of the senses of formality distinguished by MacFarlane (2000).

But how about the second combination? Saying that formality constrains validity sounds very close to defining validity in terms of formality (although it isn't *necessarily* so, as we might want to take into account non-definitional characterizations of validity as well). Broadly speaking, Tarski's project is included in the wider enterprise of explaining validity in terms of formality, for the reasons I have mentioned above. However, as we will see, MacFarlane tries to rehabilitate Tarski's account of validity, which Etchemendy had shown to be fundamentally flawed by pointing out that interpretational semantics doesn't guarantee that validity is equivalent to necessary truth preservation.

3.2. MacFarlane on formality

In this section, I introduce MacFarlane's classification of formality types. Because MacFarlane argues that one of the formality types is connected to logical normativity, his discussion of formality can be seen as preparatory to his discussion of the normativity of logic, which is the focus of this chapter.

In his doctoral dissertation, MacFarlane distinguishes three types of formality and motivates the distinction by citing the historical development of the characterization of logical validity over the past four centuries.

> To say that logic is 1-formal is to say that its norms are *constitutive* of concept use *as such* (as opposed to a particular kind of concept use).
> [...]
> To say that logic is 2-formal is to say that its characteristic notions and laws are indifferent to the particular identities of different objects. [...] Mathematically, 2-formality can be spelled out as invariance under all permutations of the domain of objects.
> To say that logic is 3-formal is to say that it abstracts entirely from the semantic content or "matter" of concepts [...]. (MacFarlane (2000): 51)

These three types of formality are independent from each other, although they can come together in certain philosophical theories. MacFarlane explains that, historically, 1- and 3- formality had been discussed long before 2-formality entered the scene. In particular, both 1- and 3-formality were already recognizable in Kant's distinction between laws of formal logic and laws of the special sciences. However, whereas philosophers after Kant continued to focus on 1-formality, 3-formality has been brought back to the philosophical arena only over the last century, when some members of the Vienna Circle started discussing logic in terms of semantic-content neutrality.

Regarding 1-formality, MacFarlane claims that it is essentially normative for thought in general, and that 1-formality norms are *categorical*. To make this point clear, it is worth mentioning a distinction which MacFarlane adopts from Kant. That these norms are categorical means that they are *universally applicable*, i.e. their application is not limited to a certain kind of concept (e.g. moral concepts) or situation. The type of norms applicable to specific kinds of concepts are called hypothetical, because they are of the form 'If condition X holds, then do Y'. As categorical norms are universally applicable, they are constitutive of thought, although this doesn't imply that the possibility of error is ruled out by these norms being constitutive of thought. These norms are not absolutely compelling in the sense that reasoners are forced to think according to them no matter what, but just in the sense that reasoners must be held responsible to the laws of logic for the assessment of their thought, MacFarlane explains.

On the other hand, 2-formality is cashed out as permutation invariance, a constraint which has become common in the logical literature especially after Tarski.[3] 2-formality avoids any reference to the norms of thought and to domain-neutrality, but represents the generality of logic as indifference to the objects' identity. Tarski's so-called interpretational semantics is clearly an example of 2-formality: Tarski defines logical validity as truth in every model, where the variation across models is taken to be a reinterpretation of the non-logical constants. In other words, logical validity and logical consequence are accounted for by Tarski in terms of truth independent from the particular identity of objects, or permutation invariance.

[3]However, MacFarlane points out that the core idea of 2-formality dates back to the nineteenth century, when Felix Klein employed the notion of invariance under a group of transformations to define different geometries. 2-formality was subsequently extended to logic (MacFarlane (2000): 57). The kernel of 2-formality presumably dates back to Bolzano as well, as Tarski admitted that his work was to be seen as continuous with Bolzano's.

What MacFarlane means by 3-formality is not as easy to grasp. 3-formality is supposed to be an absolute abstraction from semantic content, even when the content is extremely general. Whereas 2-formality abstracts from specific semantic content (i.e. the identity of objects) and maintains general semantic content (i.e. logical constants) nonetheless, 3-formality abstracts from semantic content *tout court*. According to MacFarlane, an example of the third type of formality can be found in Kant's discussion of the modes of judgment in the *First Critique*, where he considers only their *form* and abstracts completely from any semantic content that these judgments may have.[4]

Prima facie, a concept of formality (and of logicality) which abstracts from logical constants can strike one as immediately suspicious. Upon slightly more careful reflection, the advantages that such notion of formality can bring start to crop up: first of all, 3-formality escapes the modern logician's worries about determining what logical constants are. Characterizing logicality as 3-formality also seems to imply that logic doesn't provide *any* truth. In fact, whereas 2-formal logic still delivers very general truths (being concerned with general semantic content, i.e. with logical constants), 3-formal logic is supposed to abstract even from logical constants and it isn't clear how this can provide truths at all. But my concern is about the status of 3-formal logic: it seems to deliver a taxonomy of judgments, but it is hard to picture a (3-formal) logic which is *just* a taxonomy. Therefore, I imagine that 3-formal logic is also meant to be a principled way to connect judgments of a certain type with judgments of other types, as it may be the case in, say, Aristotelian syllogistics.

I find 1-formality the most interesting among the three types of formality distinguished by MacFarlane, partly because there is no clear way to connect 2-formality to our thought in a normative fashion, it being defined in terms of mathematical frameworks. On the other hand, 3-formality looks like a slightly cumbersome notion: 3-formal logic is supposed to define the conceptual (or linguistic) framework for stating facts and, as such, it seems to have a normative role, but if so, it isn't clear whether its normative role relates to 1) how our judgments or sentences should

[4]To refresh the reader's memory about the modes of judgment in Kant's *First Critique*, here is the *Table of Judgments* presented in Kant's main work.

Quantity	*Quality*	*Relation*	*Modality*
Universal	Affirmative	Categorical	Problematic
Particular	Negative	Hypothetical	Assertoric
Singular	Infinite	Disjunctive	Apodictic

be formed in order to be meaningful, 2) how they should be connected to each other to preserve truth (or validity), as in syllogistics, or 3) both.

Leaving the worries about 3-formality aside, in the rest of the chapter I will focus on MacFarlane (2004), where it is argued that formality is the source of the normativity of logic, and where the connection between formality and logic's constraining power on beliefs is cashed out in terms of bridge principles. Before turning to discuss the general framework to make sense of the normativity of logic, I want to motivate why we should embark on such an enterprise in general. To do so, I will argue against a particular anti-normativist stance in the study of human reasoning and will show that it relies on an assumption that is very hard to justify.

3.3. *Pro* normativism

Normativism with respect to the study of human thinking is the view that reasoning conforms to some extent to a normative system on the basis of which it should be assessed. On the other side, we have *descriptivism* with respect to the study of human thinking, i.e. the view that the study of reasoning should consist of the analysis of the way it actually works. It is worth pointing out that these paradigms on reasoning refer especially to the study of higher mental processing, i.e. decision making, judgment and, in a word, what we intuitively mean when we think about 'reasoning'. In fact, the literature on mental processes distinguishes between two main types of processes (this is called the *dual-process framework*), one of which is fast and undemanding and can be approximately referred to as *heuristics*, and the other is slow and demanding and corresponds to our analytic abilites, mathematical reasoning, deliberation and speculation.

Under the broad category of normativism, a corner is occupied by what Elqayam and Evans (2011) call *logicism*. Logicism understood this way is completely different from Russell and Frege's logicism about the status of mathematics. In fact, logicism with respect to human thinking is a version of normativism that takes some mental version of classical logic to be the normative system to which reasoning should conform and against which it should be evaluated. To keep things simple, let's say that any mental logic (not necessarily classical logic) will do.

I will now focus on Elqayam and Evans (2011), who argue that normativism is both harmful to and unnecessary for the study of reasoning. As a preface to the discussion of their article, I want to point out that I don't think that descriptivism and normativism with respect to the *study* of human thinking has to be seen necessarily as contrasting views. On the contrary, I think that they are wholly compatible, as they can be understood as two distinct phases of the work of the psychologist interested in the study of reasoning.

Their argument is developed around three main issues, dubbed the *arbitrariness* problem, the *is-ought inference* problem, and the *research biases* problem.

The first problem on the list addresses the situation in which the analysis of human thinking faces conflicting norms. The structure of the argument is the following: if we are to be normativist with respect to human thinking, then presumably there must be some norm that draws the line between right (or good) and wrong (or bad) reasoning. Moreover, if classical logic is the normative system human reasoning is judged against, it looks like conformity to logical laws is the norm reasoning should be assessed against. Elqayam and Evans (2011) add that, for a normativist theory to be coherent, the norm should be unique. However, when we try to make sense of the experimental results in the psychology of reasoning, it is hard to find a norm that justifies the paradigm in a coherent way. On the other hand, having more than one norm results in two problems, i.e. 1) having a conflict between norms complicates the assessment of the mental process, and 2) if the conflict is not exclusively between different norms but also between different normative systems, it results in the dilemma of how to decide which of the competing normative systems is the most appropriate to the context and therefore ought to be chosen. I will tackle the problem raised from the first point of Elqayam & Evans's argument in the later chapters of my thesis, so I postpone the discussion on it to then.

The second main point against normativism in Elqayam and Evans (2011) is meant to show that the study of human thinking in a normativist fashion is spoiled by a repeated application of the controversial *is-ought* inference. The illegitimate application of the *is-ought* inference springs from a fundamental misunderstanding of the theoretical fruitfulness of the competence theories (which the authors take to be descriptive theories): normative theories have relied on competence theories as evidential support, Elqayam & Evans argue. As it happens in linguistics, in the study of human reasoning the competence/performance level and the normative level have to be neatly distinguished, Elqayam & Evans continue. However, normativism has fallen short of doing so. In analyzing human reasoning, the descriptive role of the competence theories has been largely overlooked. In addition, Elqayam & Evans argue, normative systems, such as, say, classical logic, are pretty arbitrarily chosen to fill that role. Whatever the common habit of the normativist psychologists has been so far, I think that this analysis is flawed in many respects. In my opinion, the main problem with the argument is that it is contentious that the competence level of the analysis can really be treated as descriptive instead of as normative. Elqayam & Evans argue in favour of their thesis by using an example drawn from linguistics. In linguistics, there is a formal grammar, which fills the role of the normative system, and there is a theory that describes the language rules actually used by the speakers in everyday usage (the competence theory). Appealing to the definition of the competence theory as providing a collection of rules, the authors argue that the competence theory has a purely descriptive value.

Descriptive and normative theories differ in the questions they are developed to answer, i.e. 'what *is* ...?' or 'how *is* ...?' for the descriptive theories and 'what *ought* to be ...?' for the normative theories, and, as we expected, competence theories ask questions falling under the former kind. However, I think that an easy point can be made, which casts serious doubts on the idea that competence theories are intrinsically descriptive. Granted that they are but collections of rules governing our everyday usage of language, inferences, etc., we should be able to see that, in order to isolate such a collection of rules, we have to contrast them with the brute facts of performance, and from there we can then proceed to fill out a list of the competence rules. That is to say, to develop a competence theory, we cannot just list a collection of performance cases but we must remove what we regard as the flaws (i.e. the fallacies, errors, trivial mistakes, etc.) from them. Therefore, it looks like competence theories aren't limited to answer *what-is*-type questions and are not in fact purely descriptive as Elqayam & Evans suggest. If I am right and competence theories are normative themselves, there is no way to argue that normativism commits the naturalistic fallacy of inferring a norm from an observation.

The last problem addressed by Elqayam & Evans relates to the research biases brought about by normativism. According to the authors, adopting a normativist stance negatively affects the result of the empirical tests. In fact, assuming normativism in conducting experiments on reasoning is an instance of the *ought-is* inference, which is as illegitimate as the *is-ought* inference. However, I think that normativism shouldn't be blamed for being the exclusive carrier of the *ought-is* fallacy and of theory-laden experiments. This is a general problem that applies to experimental studies, and this point can't be considered too harmful against normativism, unless one agrees that normativism has the intrinsic feature of having theoretical biases underlying empirical research.

From the discussion of Elqayam and Evans (2011), I hope to have shown to the reader that the anti-normativist's argument is not successful. In the next section, I will delve into the normativity of logic. In particular, I will focus on the principles that bridge logical consequence and reasoning, especially in a framework developed to make sense of how logic constrains beliefs. I will highlight how some of these principles can be *prima facie* preferred over the others and I will point out that an attempt to generalize the form of the principle is not successful.

3.4. Norms for reasoning

1-formality says that logical laws are constitutive of thought *as such*. What does this amount to? To explain how logical norms are relevant to thought and to what extent, MacFarlane (2004) first distinguishes two senses of 'reasoning', i.e. a *formal* and an *informal* one. Following Harman (1986), MacFarlane defines *formal reasoning* as the reasoning one applies in drawing a conclusion from a set of premises, and *informal reasoning* as reasoned change in view. A first reply to this distinction is that the borders between informal and formal reasoning seem blurred most of the time, and therefore it might be contentious that such a distinction really exists. To prove that informal and formal reasoning are two separate modalities of thinking, the following argument is often shown:

Suppose that we have a set of premises, Γ, and that $\Gamma \vdash A$. If A is absurd or false, reasoners are rationally entitled to revise their beliefs in one or more of the premises in Γ. But if our reasoning were formal and consequently constrained by formal logic, we ought to conclude A instead of revising the premises.

I am not sure that this argument is correct, though. First of all, it seems that revising the propositions in Γ should be described as an application of Modus Tollens and of the Completeness and Deduction theorems rather than as the application of some reasoning principle independent of logical reasoning. For the time being, I will leave the question about the distinction between formal and informal reasoning open, as it is not the topic of this chapter, but I will address it later on, as this question has important relations with the way we model the structure of our mind and, more generally, with models of rationality.

For now it will be enough to say that, whereas Harman argues that there is no way to formulate plausible bridge principles between formal and informal reasoning, MacFarlane tries to articulate a spectrum of the possible bridge principles connecting the two types of reasoning, which, in turn, correspond to different ways of explaining the normativity of logic.

According to MacFarlane, the general form of the bridge principles connecting formal and informal reasoning is:

> If A, $B \models C$ then ... (normative clause).

The types of bridge principles can be distinguished on the basis of:

- the deontic operator they involve (i.e. *ought*, *is permitted*, *has reasons for*);
- the polarity of the normative clauses (are they reasons for *believing* or for *not disbelieving*?);
- the scope of the deontic operator (i.e. the deontic operator ranges over the *whole conditional*, its *consequent*, or *each of the antecedent and the consequent*).

I refer to MacFarlane (2004) for the whole list of bridge principles. For the moment, it will be sufficient to mention some of them and explain what the corresponding labels stand for. For instance, $Co+$ means that the deontic operator applies to the conclusion (C), that the deontic operator is an *ought* (o), and that the polarity is positive, so the bridge principle commits the reasoner to believing the conclusion (instead of not disbelieving it). Then, the whole bridge principle will read as

> If A, $B \models C$, then if you believe A and you believe B, you ought to believe C.

On the other hand, Bp- means that the deontic operator (which in this context is permission, p) applies to both the antecedent and the consequent of the implication (B) and the polarity is negative, so the bridge principle tells the reasoner that she is permitted not to disbelieve the conclusion. The full bridge principle is:

> If A, $B \models C$, then if you may believe A and believe B, you are permitted not to disbelieve C.

To conclude the batch of examples, let's take *Wr+*. This means that the deontic operator ('has reason for', r) ranges over the whole conditional (W) and that the polarity is positive, i.e.

> If A, $B \models C$, then if you have reason to see to it that if you believe A and you believe B, you believe C.

This taxonomy can be augmented by a whole class of principles if the antecedent of the bridge principle is taken to be 'If you *know* that A, $B \models C$', in which case the suffix '-k' is added to the principles' name.

Out of these principles, MacFarlane will consider two plausible. Co+ is the first bridge principle to be ruled out, by an argument drawn from Broome (1999). The argument goes as follows: in any logic where $A \models A$ is a theorem, the relevant instance of Co+ turns out to be 'If $A \models A$, then if you believe A, you ought to believe A', which sounds just like a self-justification of belief A, for any A. Moreover, the application of Co+ to $A \models A$ is also an instance of the the so-called *is-ought* inference. The same argument against Co+ seems to apply to all the other C-bridge principles, so MacFarlane rules out also the other C's.

As for the B-principles, MacFarlane argues that they aren't really as normative for thought as they should be. For instance, Bo+ would read as 'If A, $B \models C$, then if you ought to believe A and believe B, you ought to believe C'; however, this sounds too weak a normative principle, because 'if you ought to believe A and believe B' implies that the principle applies to those reasoners 'whose beliefs are already in order' (MacFarlane (2004): 9). In other words, the B-principles are not enforced unless the agent believes the premises that she *should* believe.

If Wo- leaves room for the possibility of revising the premises, the same should be true of Co-. Ruling out all of the C's and B's versions of the bridge principle, leaves us with six W's combinations to assess, i.e. Wo+, Wo-, Wp+, Wp-, Wr+, Wr-.

However, again, the Wp-principles seem too weak, because they state that if A, $B \models C$, you may see to it that if you believe A and believe B, you believe (or don't disbelieve) C. However, simple permission to see that if you believe the premises, you believe the conclusion doesn't really put constraints of any kind. Although this argument is used against the Wp-principles, I think that it applies to all of the bridge principles in the taxonomy involving permission as their deontic operator. In my opinion, a similar argument applies also to the bridge principles involving 'has reason to see to it'.

However, to go back to MacFarlane and cut a long story short, MacFarlane finally acknowledges Wo-, Wr+, and Wr- as the only plausible principles in the spectrum (and possibly Wr+/- as even less problematic than Wo-). Wo- reads, 'If A, $B \models C$, then you ought to see to it that if you believe A and you believe B, you do not disbelieve C'. The reason why MacFarlane thinks that Wo- works better than Wo+ is that, including the positive polarity in the consequent ('you believe C'), Wo+ seems to assume that the rational agent's beliefs are closed under logical consequence.

The only concern that one might have about Wr+ and Wr- is that they don't impose strong constraints on belief, but rather recommendations on what to believe. The Wr principles have weaker normative power than that of Wo-, nonetheless they have a slightly stronger power than permissions. Being permitted to believe the consequence upon believing the premises of an argument says that one *can* conclude C from A and B, but it doesn't properly *recommend* it. On the other hand, having a reason to believe the conclusion upon believing the premises of an argument is enough to recommend that one believes the conclusion.

3.4.1 The subjective acknowledgment turn

So far, so good. What happens next? MacFarlane revises the bridge principles in light of further considerations on transparency and formality. Let's start with formality. MacFarlane (2004) points out that 'the normativity of logic has its source not in the formal validity of inferences, but in the formal validity of inference schemata' (*ibid.*: 22). The point is made to exclude material validity from the discussion on bridge principles, and I won't argue against this for the moment.

In addition, MacFarlane claims that, for a bridge principle to be fully enforced, logical validity has to be transparent. In other words, for a particular occurrence of the bridge principle regarding the inference A, $B \models C$ to have an effective normative power over an agent, the agent must recognize that A, $B \models C$ is an instance of a valid inference schema. For instance, transparency fails when two coreferring terms occur in an inference without the agent being aware that the two terms denote the same object, as in the famous 'Hesperus is Phosphorus' case.[5] In such a case, MacFarlane argues, the agent is at least excused for violating the bridge principle to some extent (MacFarlane (2004): 21, footnote 23). This calls for a revision of the form of the bridge principle in such a way that the principle makes explicit the role of inference schemata in the normative power of logical validity and avoids holding an agent responsible for violating the bridge principle when cases similar to 'Hesperus is Phosphorus' occur. The revised version of the principle form reads:

> If [you know that] the schema S is formally valid and you apprehend the inference A, B / C as an instance of S, then (normative claim about believing A, B, and C). (MacFarlane (2004): 22)

However, is this revised version of the principle really an improvement? Consider the ways in which an agent can fail to recognize an inference, I, as an instance of an inferential schema, S:

1. The agent knows that S is a valid inference schema but fails to apprehend I as an instance of S.

2. The agent doesn't know that S is a valid inference schema.

Then, if an agent fails to recognise I as an instance of S, that might happen because the agent doesn't know that S is a valid inference schema. This would mean that the revised bridge principle takes into account not only the cases in which logical validity is not transparent, but also the agent's logical competence or logical knowledge. In other words, the principle enables the agents to be freer to believe what they like the more logically ignorant they are. However, this very point seems to be inconsistent with a comment that MacFarlane makes on the -k bridge principles.[6] The issue is labelled 'Priority Question' by MacFarlane, and is presented in the following way:

---

[5] In such a case, an inference like

> Hesperus is the first star to appear in the evening.
> Phosphorus is the last star to disappear in the morning.
> Therefore, some star is the first to appear in the evening and the last to disappear in the evening.

is valid in Kaplan's sense (see Kaplan (1978)).

[6] The -k bridge principles are those of the form, 'If *you know that* A, $B \models C$, then (normative claim)'.

> According to the -k variants, we are subject to logical norms only in so far as we have logical knowledge. The more ignorant we are of what follows from what, the freer we are to believe whatever we please [...]. (*Ibid.*: 12)

As we have seen, the same consideration holds also for the non-k bridge principle. Therefore, if we are to exclude material validity from our normative account, we have at least to get rid of the subjective acknowledgment bit of the new bridge principle. At that point, it can be questioned whether the shift from a class of bridge principles to another is really worthy, though, and we might just decide to be content with the old principles. I will temporarily leave aside the question about the general framework of the bridge principles, and will turn to the deontic operator involved in the bridge principles.

3.5. Analysis of the deontic operator

3.5.1 Senses of *ought*: suggestions from the study of rationality

Five senses of *ought* can be found in the literature on rationality. Given that we are concerned with constraints (and specifically, with obligations), I think that it would be helpful to specify in what sense *ought* is used in the context of logic. The first three senses in the list below (i.e. *functional, directive*, and *evaluative ought*) are drawn from Thomson (2008) and Elqayam and Evans (2011), whereas I have taken the distinction between *subjective* and *objective ought's* from Kolodny and MacFarlane (2010). Later on, I will address the question of what the interplay between these senses of *ought* is.

(1) *Functional ought.* This is an impersonal *ought* that applies to objects and refers to their proper function, as in 'Heart ought to pump the blood'.
(2) *Directive ought.* This is an instrumental *ought* that applies to agents, as in 'To get to the entrance door, you ought to walk behind the building'.
(3) *Evaluative ought.* This is a "principled" *ought* that applies to both objects and agents, such as in 'We ought not to kill animals'.

(4) *Subjective ought.* This is relative to what you should do (or believe) given only the information available to you.

(5) *Objective ought.* This is relative to what you should do (or believe) given all the information.

Apart from the functional sense of *ought*, which is impersonal, we can ask whether directive and evaluative *ought's* can come in subjective/objective versions. It is pretty clear that the directive *ought* can take subjective or objective sense, as in the example of the miners trapped in a shaft (Kolodny and MacFarlane (2010), although the example I use here is slightly modified). I will summarize the example to give a taste of what the subjective and objective readings of the directive *ought* amount to.

Ten miners are trapped in one of the two shafts of a mine, call them A and B, but we don't know in which one. A big quantity of water is expected to hit the shafts and the rescuers have to decide where to deviate the water to. If the entrance of one of the shafts is blocked, the water will fill the other shaft to the ceiling and all the miners will die; if neither of the shafts is blocked, the water will flow into both shafts, filling each of them halfway up, and only the shortest of the miners is likely to drown. Not knowing in which of the shafts the trapped miners are, the rescue team *ought* to block neither shaft, so that nine out of ten miners will survive. But suppose that someone, say, Moira, has a very sophisticated machine that detects heartbeats at very long distances and that this machine detects that there is someone in shaft B and no one in shaft A. Moira goes to the rescue team and advises them 'I know where the miners are. You *ought* to block shaft B'.

The *ought's* in question in making decisions about which shaft has to be blocked are not of the evaluative type. In fact, there is no principle involved in blocking a shaft or the other, or neither, *per se*. These *ought's* are directive, as they refer to actions that are instrumental to save the miners. On the other hand, in the trapped miners' story, an evaluative *ought* could be expressed as 'We ought to save the highest number of people'. Whereas it is pretty uncontroversial that the directive *ought's* can be either subjective or objective, it still can be asked whether the evaluative *ought's* can, too. *Prima facie*, it doesn't look like the quantity of information available to the agent makes much difference in the context of evaluation. However, we can be easily distracted by a different sense of 'objective' and 'subjective' here, and be subsequently led to think that there is in fact something like a subjective sense of the evaluative *ought*. For instance, suppose that Irene and Ida are discussing Irene's decision to become a vegetarian. Irene says: 'You *ought* to quit eating animals'. Ida: 'I can't see why. I think that you *ought* not to become a vegetarian'.

Although the *ought's* in the dialogue above are evaluative and they are relative to two different points of view (Irene's and Ida's, in the example), it doesn't look like they are subjective *ought's just because* they are expressed by different subjects and are relative to their individual views. We might report the conversation as: 'According to Irene, you ought to become vegetarian and according to Ida, you shouldn't', but this has nothing to do with the amount of available information that Irene and Ida have at that point.

For now, let's say that the constraint that Wo- imposes on beliefs could be seen as a mix of the directive and of the evaluative *ought*. In fact, if the obligation involved in the normativity of logic were purely evaluative, it would be hard to explain how advocates of different logics can have a genuine debate about logic and not just talk past each other, as in Irene and Ida's case. In the last chapter, I will argue that the normativity of logic is in fact much weaker than the prescription imposed by the *ought's* would suggest, but for now I want to focus on the analysis of those bridge principles that came out as winning the race. The next section is therefore devoted to a possible analysis of the deontic operator 'has reason to'.

3.5.2 More deontic operators

The study of the normativity of logic on thought is not exhausted by the discussion of the bridge principles, though. To make clear what is the exact meaning of the principles, it is not enough to rely on the intuitive understanding of the deontic operators involved. The deontic operators, as they have been used in MacFarlane (2004), are addressed by an extensive literature, and they constitute an object of investigation on their own. The three deontic operators considered in drawing the bridge principles are *is permitted to*, *ought* and *has reason to*. The definitions of *is permitted to* and *ought* are such that we can take one of the operators as basic (e.g. *ought*) and define the other operator in terms of it, analogously to what we do with $\Diamond$ and $\Box$. In fact, the permissibility operator (let's call it P) is usually defined as follows: $P\varphi =^{df} \neg O \neg \varphi$. On the other hand, the obligation operator, O, is such that $O\varphi$ is equivalent to $\neg P \neg \varphi$. The third operator, *has reason to* (let's call it R), is taken from Broome (1999), where the difference between reasons and normative requirements is spelled out. The distinctive feature of *has reason to* is that an agent can have conflicting reasons, i.e. reasons for φ and reasons for $\neg\varphi$, whereas conflicting obligations are usually not admitted.

Chellas (1974) develops a logic for the monadic (and the dyadic) deontic operator and two systems for each of the operators. I will now focus on the monadic deontic logic introduced by Chellas and argue that, despite the fact that the two systems (D and D*) presented in Chellas (1974) are both meant to be systems for the operator O, one of them is also appropriate for a different operator, i.e. R, discussed in Broome (1999). I present the main features of D and D* below.

System D* (*standard monadic deontic logic* in Chellas (1974)):

OA is true at w if A is true at w's deontic alternatives. Every world, w, has a related set of possible worlds, *DeAl,* that are w's deontic alternatives. Therefore, $\nu_w(OA) = 1$ if $\forall w' : w' \in DeAl_w$, $\nu_{w'}(A) = 1$.[7]

(ROM) $\frac{A \to B}{OA \to OB}$

(Obligation is closed under logical implication)

(OD) $\neg O\bot$

(Nothing impossible is obligatory. Transposed: Ought implies Can)

(ON) $O\top$

(Obligations exist at every possible world)

(OK) $(OA \wedge OB) \to O(A \wedge B)$

---

[7] Another way to cash out the definition of O is to specify a relation, R, meaning 'is a deontic alternative to ...', such that $\nu_w(OA) = 1$ if $\forall w' : w' \in W$, such that w'Rw, $\nu_{w'}(A) = 1$.

(Obligation is closed under adjunction. Chellas (1974) argues that (OK) amounts to closing obligation under logical consequence - it states 'that a proposition is obligatory if it is a consequence of obligatory propositions'[8])

D* doesn't allow any conflicting obligations. In fact, take (OD) and consider a contradiction, e.g. $A \wedge \neg A$. By (OD), $\neg O(A \wedge \neg A)$. Therefore, by transposition of (OK), $\neg(OA \wedge O\neg A)$.

These axioms conform to the meaning that is usually given to the deontic operator O. However, removing (OK) from our deontic logic results in $\neg(OA \wedge O\neg A)$ not being derivable anymore in the system. The new system is what Chellas (1974) calls D, or *minimal monadic deontic logic*. To be more precise, in D neither (OK) nor (ON) hold. The fact that (ON) doesn't hold means that there are possible worlds without any obligation. However, as D and D* provide different axioms for the same operator, i.e. the monadic *ought*, they ultimately leave to the reader the burden of figuring out which *ought* she is talking about each time - the strict obligation of D* or the weaker obligation of D?

I suggest that we retain both the logics D* and D, interpreting the deontic operators used in each system instead. I will therefore ask the reader to follow me in a short detour through the deontic operators, to identify what operator best suits each of the systems discussed above.

As I mentioned, Broome (1999) distinguishes reasons from normative requirements. In doing so, he argues that reasons impose on the agent a constraint which is weaker than the demand imposed by normative requirements. To this purpose, Broome proposes that we use two deontic operators, R and O, where R (i.e. 'has reason to ...') is weaker than O (i.e. *ought*). The distinguishing feature of R is that an agent is allowed to have conflicting reasons, namely, reasons for A and reasons for *non-A*, as *having reasons for* is *pro tanto*, i.e. it is not as constraining as *ought* is. At the same time, if you have a reason for A and no reason for *non-A*, you ought to A, therefore $(R(A) \wedge \neg R(\neg A)) \rightarrow O(A)$.

It seems plausible that D is suitable for R, as it allows for conflicting reasons. Therefore, I suggest that we replace O with R in system D, such that

(RRM) $\frac{A \rightarrow B}{RA \rightarrow RB}$

[8]Chellas (1974): 23-24.

(RD) $\neg R\bot$

On the other hand, as long as we don't allow conflicting obligations (that is, at least when the objective *ought* is in play), we can be quite happy with D*. However, if obligations are closed under logical consequence, this might be a feature that doesn't match with the normative requirements that logic imposes on thought, as the *ought* involved in the bridge principles seem to be the subjective *ought* rather than the objective *ought*.

So far, I have taken into consideration only monadic deontic operators. Now I want to spend a few words on the reason why I didn't take conditional obligations into account in my discussion of the deontic operator suitable for the bridge principles. Suppose that we want to provide a version of the bridge principles that employs a dyadic deontic operator, such as $O(A|B)$, to be read as 'You ought to believe A, given that you believe B'. A bridge principle with conditional obligation would read as

> If A, $B \models C$, then you ought to believe C, given that you believe A and believe B.

Now, let's consider the instantiation of such a bridge principle for $A \models A$. That would read as

> If $A \models A$, then you ought to believe A, given that you believe A,

which is as unconvincing as the C-versions of the bridge principles are (for the same reason). Then, it looks like conditional obligation is not a plausible candidate for the bridge principles, and we should stick to monadic deontic operators.

3.6. Conclusion

We have seen that the bridge principles are an attempt to explain the normative power of formal reasoning over informal reasoning. Among the options presented by MacFarlane, only three can be considered plausible (at least *prima facie*), and I argued that trying to generalize the principles by appealing to inferential schemata faces an important problem.

I distinguished five different senses of *ought* (of which, only four are relevant to the normativity of logic), and suggested that the *ought* involved in the bridge principles is a directive obligation, which leaves room for genuine debate about logic.

Finally, I focused on the operator *has reason to*, which first appeared in Broome (1999) (as far as I know) and was used in the bridge principles Wr+ and Wr-, and suggested that it can be made sense of by looking at Chellas's minimal deontic logic.

The discussion in this chapter lays the ground for the analysis of the normativity of logic that takes place in the sixth chapter. That is why I haven't taken a stance on which bridge principles are to be preferred over the others. Although I agree with most of the initial considerations that MacFarlane does about the scope of the deontic operators and the types of operators (e.g. ruling out permissions), the discussion in this chapter is meant to be propaedeutic to the rest of the thesis.

CHAPTER 4

Minimal rationality

Abstract. In this chapter, I address two theories of rationality that have changed profoundly the literature on reasoning and rationality. The two approaches are presented in Harman (1986) and Cherniak (1986) and have several elements in common. First of all, both theories focus on the study of rationality in real agents; both approaches take into account the cognitive, memory and temporal limitations of human agents. Discussing these theories of rationality will help lay the ground for discussing the normative role of logic on thought in real agents, which is the ultimate goal of my thesis. The last section of the chapter is devoted to discuss Peirce's account of *logica docens* and *logica utens*, which is beneficial for putting into perspective the accounts of rationality discussed in the earlier sections of the chapter.

Introduction

Before we investigate the normative status of logic, we need to qualify *what* it is that we are judging logic against. In other words, we take logic to have a normative status because it provides norms that should help us maximize rationality, but what kind of rationality should we consider? As our concern is with ordinary agents, this chapter describes some theories of rationality that have been developed to capture the sort of rationality we actually expect to find in our and other people's beliefs and behaviours. For this reason, it is important to stress that belief revision is relevant to logic. This would perhaps strike some logicians as an undesired piece of psychologism but rather it is crucial for logic to take belief revision seriously. By this I don't mean that logic ought to represent or simulate belief revision, although logics of belief revision have already been developed, e.g. AGM framework by Alchourron and Makinson (1985). Instead, belief revision is relevant to logic as a scientific theory that evolves over time and this looks like a good reason for logic to care about change in belief. Furthermore, belief revision is something logic should care about because we want logic to give us some constraints, directions, or somehow to guide us when we add a new belief to our pre-existing set of beliefs. In this sense, we want to spell out what logic can do to guarantee that our beliefs be rational, which is so important to us because we decide and act upon them.

Belief change is not necessarily synonymous with rational belief change, though. For example, take Dave's particular belief, b, that plane trails are caused by condensed water vapour resulting from the exhaust of aircraft engines. Now imagine that Dave reads a leaflet published by an alleged scientific association explaining that some of those trails are not actually made of condensed water vapour but rather are chemical agents sprayed by planes to harm people. Dave finds this theory convincing and comes to believe that plane trails are actually chemical trails aimed at intoxicating people. Although Dave might have come to believe what the leaflet says because he finds the argument particularly compelling or cogent, from an objective point of view, this is an obvious case of bad belief revision and, in fact, its outcome is less rational than the starting belief. Therefore, there must be something involved in belief change which turns a belief revision into a rational revision, i.e. good/reliable justification. Reliability and justification are objects of a massive investigation in epistemology which is beyond the purposes of this chapter, so I will leave the issue aside. For the time being, let's say that we want to consider logic(s) as a set of beliefs which can be revised according to some criteria, the investigation of which has been extensively undertaken in the works I am going to discuss in the following sections.

For almost a century, the issue of logic being relevant to rationality hadn't been seriously put into question. Frege argued that if logic (recall that by 'logic' he meant some sort of ancestor of contemporary classical logic) is the study of validity in deductive arguments, it has to be rational and nothing is more tightily connected to rationality than logic itself. In fact, according to Frege, logic has a special status among other sciences, and it is no doubt the science conforming to rationality itself at the highest levels because rationality is constituted by logic. If this can be paraphrased as 'logic is rational because rationality is logical', Frege's position on the connection between logic and rationality can hardly be seen as helpful in clarifying the matter.

On the other hand, trying to give a comprehensive definition of 'logic' and trying to analyze its interplay with rationality raises many questions, especially because it is challenging to define the rather vague concept of rationality. We might even find it difficult to get a unitary and non-subject-specific articulation of rationality. As for logic, the situation is perhaps less ambiguous as we all know that it is defined as the systematic study of the laws of valid inference, but at the same time, as we have seen in the first chapter, it is controversial how validity itself should be defined; in the first chapter, we have seen that a definition of validity can be given in - at least - three different ways, i.e. model-theoretically, proof-theoretically and non-definitionally (e.g. characterizing validity in terms of degrees of belief, as suggested in Field (2009)).

We have already discussed the contrasting positions on logical form and on the nature of validity, so now I will rather move on to discuss some views of rationality. In this chapter, I will first describe two approaches to rationality sharing the common thesis (labelled '(*ILR*)' in the following sections) that logic is not specially relevant to rationality; in the later chapters I will argue whether *ILR* is really justified.

4.1. Is deductive reasoning a chimaera?

As anticipated above, I am going to present two approaches to rationality which share a common thesis about the relevance of logic to rationality. These theories also have in common an important assumption on the nature of rational agents, as we will see in due course. The core works that I am going to discuss here are Harman (1986) and Cherniak (1986), which represent important achievements in the study of what could be called *minimal rationality*.

Harman (1986) argues that logic is neither a psychological nor a normative theory, which means that logic doesn't describe how people actually reason nor is it an adequate theory about how people ought to reason in order to be rational. To support this point, Harman first distinguishes between *reasoned change in view* (or reasoning) and *argument* (or proof). According to his view, deductive principles, which are employed in arguments, are of no help when it comes to revising our beliefs. To see his point, suppose that $\Gamma \vdash A$, where Γ is a set of beliefs and A is any proposition logically derivable from Γ. Now, if an agent discovers that A is not the case, then she will realise that something must be revised in Γ. Therefore, following logical principles can lead an agent to acknowledge that a revision process in her beliefs must be undertaken, but logic can't (and isn't supposed to) suggest how to revise beliefs.

In order to stress the difference between reasoning and proofs, Harman identifies some principles of reasoning, i.e. principles according to which rational agents revise their beliefs. The main principles of reasoning listed in Harman (1986) are:

- (*Clutter Avoidance*): One ought not to clutter one's mind with irrelevancies.
- (*Coherence*): One ought to revise one's beliefs so to avoid inconsistencies among them and to maximize the explanatory power of each on the others.
- (*Conservatism*): In revising one's beliefs, one ought to minimize the change.

As it is apparent from the list, the principles of reasoning are not principles of logic on Harman's account. Moreover, as we have seen above, Harman gives different meanings to *argument* and *reasoning* and goes on to state that, consequently, deductive argument and deductive reasoning shouldn't be identified, nor should inductive argument and inductive reasoning either. In line with his previous statements about the nature of reasoning and argument, the author claims that deductive *reasoning* simply doesn't exist, whereas inductive reasoning does. On the other hand, it is contentious that agents can produce inductive *arguments*, whereas one can certainly produce deductive arguments. To see why Harman claims that deductive reasoning doesn't exist, recall that (on Harman's account) principles of reasoning constrain how beliefs ought to be revised, whereas deductive principles provide reasons for revising one's beliefs at best.

Before we continue discussing Harman's theory of rationality, it's worth saying more about the approach to logic in Harman (1986). We have seen that Harman takes arguments (and not reasoning) to be the proper domain of logic. Harman (unpublished) denies that logic is the science of what inference forms preserve truth necessarily. This claim can be made sense of in light of the distinction drawn between inference and implication presented in Harman (1986). A good deal of Harman's first chapter is devoted to defending the view that *inference* doesn't match up with *logical implication*, arguing that the distinction between implication and inference is parallel to the dichotomy argument-reasoning. On Harman's account, implication links adjacent steps of an argument, whereas inference plays a role in reasoning and doesn't follow logical principles. Harman argues that the main point showing that implication and inference are not intersubstitutable is *cumulativity*: whereas implications are cumulative, inferences are not. In other words, Harman states that logical implication is monotonic, whereas our reasoning seems to be dynamic and inferences seem to be non-monotonic. The obvious question here is whether the same difference would persist if instead of implication in, say, classical logic, we considered implication in non-monotonic logics. Harman dismisses this option very quickly, pointing out that

> [...] although this terminology emphasizes the noncumulative character of reasoned revision, it is also potentially misleading in calling the ordinary sort of proof or argument "monotonic reasoning", because proof or argument is not of the same category as reasoned revision. (Harman (1986): 4)

However, this reply seems to beg the question, because Harman is precisely trying to support the very existence of a distinction between argument and reasoning through the distinction between implication and inference here.

Let's pause once again on Harman's position on deductive reasoning (or better, on its nonexistence). In order to show that reasoning doesn't follow deductive principles and, in turn, that deductive reasoning doesn't exist, Harman stresses that

a) logical principles hold in any case and don't admit exceptions, whereas not in every case we are inclined to infer, say, Q from P and *if* P *then* Q.

b) when reasoning, we are not usually inclined to draw all the conclusions that an application of the logical law in question would allow us to draw, e.g. from a contradiction P *and not* P we don't feel compelled to draw anything (in fact, we usually don't conclude *any* proposition whatsoever from a contradiction). We avoid applying instances of *ex falso quodlibet* (EFQ) not to clutter our minds with irrelevant information, Harman concludes.

The above two points display a clash between two competing accounts of rationality. On one side, there is a well-entrenched model revolving around an ideal agent possessing flawless and infinite cognitive, deductive and mnemonic abilities. I will label this theory *Ideal Agent Rationality*, or *(IAR)*.

On the other side, there is an emerging model of rationality which rejects ideal agents and is centred upon human agents instead, taking into account real agents' limitations in terms of cognitive and deductive abilities, information retrieval, memory capacity. I will label this theory *Minimal Rationality*, or *(MR)*.[1]

Point *b)* addresses a fundamental principle of (IAR), i.e. the

(*Logical Closure Principle*): a set of beliefs, Γ, of an agent, S, ought to be closed under logical consequence, i.e. for any proposition, A, such that if $\Gamma \vdash A$, S ought to believe A,

[1] *Minimal Rationality* is the term that Cherniak (1986) uses in his work.

and denies that Logical Closure is a crucial principle of rationality. On the contrary, rational agents would better violate Logical Closure not to clutter their minds with irrelevancies, according to (Clutter Avoidance). Harman stresses that it is real agents' limitations in terms of mental abilities and memory storage capacity to make it impossible to comply with Logical Closure.

However, I'm not sure that the points made by Harman are conclusive arguments to support the view that reasoning and argument and, in turn, implication and inference are really different. I don't mean that argument and reasoning are equivalent, but rather that Harman's argument doesn't make the right moves to dismiss the option that they are tightily connected.

To see why, let's go back again to point *b).* The remark in *b)* would succeed in proving that logical and reasoning principles have nothing to do with each other if Clutter Avoidance were an exclusive feature of *reasoning* (I am referring to reasoning as *reasoned change in view*, as Harman puts it). However, it seems to me that the same applies also to proofs and arguments: it is not the case that, just because we *can* go on drawing conclusions from certain premises in a proof, we actually do that if these are not relevant to what we intend to prove. Therefore, Harman must mean that principles of reasoning are metaprinciples. To make the difference between reasoning and logical principles clear, let's reflect on what happens when we try to prove some conclusion, C: we apply logical rules in drawing a line from another, e.g. in natural deduction, from a set of premises, Γ, to a set of intermediate steps, Δ, to the conclusion, C, but also apply a proof strategy to get the conclusion we want, e.g. we stick to a certain order in applying inference rules to keep the proof as short as possible, and this strategy might follow general principles of reasoning which are distinct from the logical rules.

It seems that, according to Harman, in this respect it wouldn't make much difference if we replaced 'logical rules' with 'chess rules' and 'proof strategy' with 'game strategy' in the sentence above: that is why talking about deductive reasoning is as meaningless as talking about chess reasoning. Leaving chess aside and going back to philosophy talk, Harman's description of principles of reasoning makes them look close to what other philosophers have called 'epistemic norms'. Harman argues that it might well be the case that we fail to apply (or consciously suspend) rules in our reasoning, as stated in point *a)* above but if the principles of reasoning were equivalent to logical principles, we should apply the reasoning principles without exception.

This is not the case and therefore deductive principles can't also be principles of reasoning. Consequently, whereas arguments can be valid or not, reasoning cannot be valid nor invalid. In other words, Harman is arguing that because rational agents can dispense from applying logical laws, the principles of logic can't be principles of reasoning, which is similar to saying that chess rules can't be game strategy principles *because* we may fail to move the rook correctly.

The overall impression is that Harman is not proving that there is a difference between reasoning and argument and between implication and inference, but he rather assumes that such differences are in place. This makes the distinctions remain ultimately not well-defined. Moreover, the mutual appeal of each distinction to explain the other doesn't help much either.

However, to give additional support to his view, Harman proceeds to list four possible ways in which reasoning can go wrong:

> 1. One might start with false beliefs and by reasoning be led into further errors.
> 2. One might reach a conclusion that is perfectly "reasonable", even though it happens to be mistaken.
> 3. One can be careless or inattentive; one can forget about a relevant consideration or fail to give it sufficient weight; one can make mistakes in long division; one can fail to see something, to remember something, to attend carefully; and so on.
> 4. One can revise one's view in accordance with an incorrect rule of revision, thereby violating the correct rules.
> (Harman (1986): 7)

Among these, Harman argues that only the last two can really be regarded as reasoning mistakes. However, given that Harman argued that principles of reasoning are principles of reasoned belief revision, it looks like the only proper error of reasoning is only 4. In fact, Harman claims, errors in group 3 can be attributed to lack of reflection, but it is not clear whether Harman wants errors in group 3 to fall under the category of reasoning mistakes or not. That obviously depends on what Harman exactly means by 'reasoning'. If we take principles of reasoning to mean principles of rational belief revision and consequently reasoning as rational change of belief, then errors in group 3 don't seem to fit in the family of reasoning mistakes. However, Harman doesn't deny that they are reasoning mistakes (as in fact they seem intuitively), but it is hard to see a view on reasoning which classifies group 3 as errors of reasoning with the meaning of 'reasoning' given by Harman and discussed above.

To sum up, according to the distinction between argument and reasoning presented in Chapter 1 of Harman (1986), on one side we have proofs or arguments, which are properly the domain of logical principles and validity and on the other side we have reasoning, which is reasoned change in view, and its principles, which look quite close to epistemic norms and ultimately metaprinciples. I don't think that Harman's argument does prove that such a distinction is in place, but we could just assume that such a distinction is an intuitive way to define different mental processes, one of which is step-by-step and cognitively demanding, while the other is intuition-based.

4.2. Which logic?

The purpose of this section is to argue that (ILR) needs to be qualified. Before I do that, I clarify what kind of logic is the object of (ILR), the thesis that logic is not specially relevant to reasoning. In his examples, Harman refers clearly to first-order classical logic, but he means some language elements, such as the truth predicate, to be included in the logic, so what Harman refers to seems to be a sort of increased classical logic which includes also some strictly semantic elements. For example, Appendix A of Harman (1986) includes the following argument:

> X is part of Y
> Y is part of Z
> Therefore, X is part of Z

and it is argued that it might be the case that we can consider it a logical argument.

Harman explains that to identify logical constants, we often apply the *grammaticality criterion*, briefly described as:

> [...] logical constants in a natural language are grammatically distinctive in always being members of small, closed logical classes of terms.[2]

[2]This passage refers to Harman (1982).

However, Harman suggests that we drop the grammaticality criterion to take the truth predicate into account as a logical constant. In fact, in order to have instances of the Truth Schema in the set of our logical formulas, we want truth to be included in the logical vocabulary, despite the fact that the truth predicate doesn't display any particular difference with respect to other predicates and adjectives that behave very similarly in 'that'-clauses, such as 'is probable' and 'believe'. Harman tries to support the thesis that Tarski's biconditional should be part of our logic to prove that only logical implication and inconsistency are psychologically 'immediate', i.e. in, say, inferring A from $A \wedge B$, no intermediate steps are necessary from a psychological point of view. However, Harman argues, this doesn't mean that logic has any privileged relation with reasoning.

It would be interesting to understand who the addressees of Harman's arguments are. From the discussion in *Change in View*, the Artificial Intelligence programme looks like a plausible candidate and perhaps it is what Harman has in mind when he distinguishes between inference and implication and argues that, specifically, non-monotonic logics[3] and, generally, logics which attempt to model human reasoning don't model how human agents *infer*. His point in raising questions about logics involved in AI concerns the impossibility of (presumably) dynamic or epistemic logics to ultimately model human reasoning, because reasoning escapes the structure of logical implication. We have seen the reasons why Harman argues this, so I am now going to make some additional remarks about Harman's claim that logic is not specially relevant to reasoning and then I move on to a similar view on rationality which displays the crucial features of (MR) (like Harman (1986) does), i.e. the thesis that (IAR) isn't a satisfactory model of rationality.

I want to conclude this section raising some questions about the plausibility of (ILR). The main remark about (ILR) is that it is not clear what it is to be 'specially relevant to reasoning'. My best bet is that this special relevance consists in *putting constraints* on rationality, as reasoning is defined as rational change in view. But what constraints are we talking about? Does (ILR) mean something like 'Logical entailment imposes obligations on what we should believe in order for us to think rationally' or does it mean something like 'Logical entailment gives us recommendations on what we should believe in order for us to think rationally'? The two options

[3] Or what we would call 'dynamic logics' nowadays.

are quite different, and it will be especially clear that they are in the last chapter of the thesis.

What Harman's argument really manages to defeat is the view that the application of inferential rules *always* results in epistemic rationality. In other words, Harman argues that (i) classically valid inferential rules are not applied systematically by human agents, and (ii) even though human agents don't apply those rules all the time, their reasoning can be deemed rational nonetheless.[4]

The claim that logic is relevant to rationality doesn't commit us to apply all the classically valid inferences, nor to stick painstakingly to predicate logic all day. What makes the difference in proving or disproving (ILR) is, as I mentioned, the normative power that we attribute to logic. There is a big difference between expecting logic to impose obligations on thought in order for it to be rational and expecting logic to guide thought.

Besides that, it is worth highlighting that Harman's work is revolutionary for the study of the normativity of logic in two senses: first, Harman gives philosophical relevance to a realistic model of rationality that takes into account the limitations of real agents; second, Harman acknowledges that logical laws and epistemic norms sometimes conflict, and that rational thought doesn't depend necessarily on following the former.

What I have shown is that the thesis that logic is not specially relevant to reasoning is in need of qualification. Firstly, I have clarified what logic Harman is concerned with in his *Change in View*. Secondly, I have distinguished two senses in which (ILR) could be read, that will be the object of discussion in the last chapter of my thesis.

I will now turn to another account of rationality, i.e. Cherniak's minimal rationality account, and I will highlight the common ideas between Harman's theory of rationality and the minimal rationality approach.

4.3. Minimal rationality

Not very dissimilar to Harman in spirit is Cherniak (1986), where rationality and the constraints it imposes are discussed. What is at stake in Cherniak's theory is

[4] In the next chapter, I will discuss experimental results in the psychology of reasoning that confirm (i), and in the sixth chapter, I will argue for a claim very close in spirit to (ii).

the foundation of our theories about other people's cognitive states. How do we attribute intentions and beliefs to ohers? Trying to answer this question, Cherniak embarks in an inquiry about human limitations of cognition, memory and deductive abilities. Cherniak wants to show that a model of rationality based on ideal agents with impeccable cognitive capacities and logical omniscience wouldn't help us much when it comes to attributing intentions and beliefs to agents or when it comes to attributing rationality to agents' beliefs and actions.[5] What we can predict about others' mental beliefs is based on our implicit acknowledgement of certain systematic shortcomings in deductive tasks, information retrieval and memory tasks.

Due to these limits, humans have developed a pretty reliable heuristics which enables them to draw conclusions and make decisions privileging quickness over information and deductive accuracy. The discrepancy between idealized cognitive and deductive abilities and the "quick and dirty" decision-making process has motivated the development of a model of rationality alternative to the (IAR). In much discussion on rationality, decision-making and game theory, ideal agents supplied with infinite deductive and memory powers have been considered a basic presupposition. Among these agents' deductive powers, the most notable ones are Logical Closure (which I mentioned above) and Belief Consistency. Logical Closure and Belief Consistency are interconnected in an important sense: if an agent S has a set of beliefs, Γ, which is closed under logical consequence, then S has also the ability to detect all the inconsistencies in Γ and to remove them. Cherniak argues that an alternative (weaker) ideal agent model can be found in the literature, e.g. in Hintikka (1962, 1970), where the idealized agent's inferential abilities conform to what has been summarized by Cherniak as *ideal inference condition:*

> If A has a particular belief-desire set, A would make all and only sound inferences from the belief set that are apparently appropriate. (Cherniak (1986): 13)

This condition would be equivalent to:

> (i) A would select all and only those inferences to make from the beliefs that are apparently appropriate for A to make.

[5]Cherniak presupposes a *holistic* approach to beliefs, desires and meanings, where these are deemed as interdependent.

> (ii) *A* would successfully perform all and only those inferences. (*Ibid.*: 13)

Cherniak dubs (i) *ideal heuristic requirement* and (ii) *ideal deducing requirement*, and points out that in practice it is never the case that these conditions are met by real agents and that human agents follow them in attributing rationality to other agents' beliefs. Agents conform to a minimal requirement on inferences:

> If *A* has a particular belief-desire set, *A* would make some, but not necessarily all, of the sound inferences from the belief set that are apparently appropriate. (*Ibid.*: 10)

In other words, of all the possible inferences which real agents might perform, only some are actually picked out in virtue of being appropriate to the agent's goal and, of these, only a subset is performed. Heuristics tells us what inferences we ought to perform to accomplish our target but doesn't explain why we select a particular subset of the appropriate inferences. Cherniak offers a theory to account for this selection, labelled as 'theory of *feasible inferences*'. According to this theory, it would be unrealistic to deem all the inferences as *maximally feasible*, i.e. always accomplished, or not feasible at all, i.e. never accomplished, because real agents tend to fall in between. Accordingly, Cherniak suggests that inferences can be ranked on the basis of a feasibility ordering which applies to agents in a temporally and contextually relative fashion. In other words, the difficulty that real agents encounter when drawing inferences may vary over time and situations and it is difficult to identify a batch of inferences which are unanimously easy to accomplish.

According to Cherniak, the impossibility of drawing a distinction which enables agents to attribute rationality a priori can be swallowed after one realizes that ideal deductive abilities would be not only inappropriate to explain rationality, but even detrimental to it. Cherniak initially takes first-order classical logic as the set of inference rules and axioms to run his argument against (ILR), arguing that whereas ideal agents would keep applying classical logic rules regardless of context and regardless of relevance of the conclusions to the context, real agents would select a subset of the inferential rules and would apply them in a context-sensitive way. Then, Cherniak argues that the argument against (ILR) holds even if classical inference rules are replaced by non-classical ones.

Besides the common focus on real agents, Cherniak and Harman's arguments are similar with regard to their approach to logic: both Harman's theory and Cherniak's minimal account of rationality revolve around the assumption that 'logic' is coextensive with '(classically) valid inferential rules'. For the sake of simplicity, in the next chapters I will also take the example of classical logic as representative in my discussion, but I am aware that sampling the class of logic more broadly would make the analysis more exhaustive.

A different perspective on logic is given by Peirce, who distinguishes different uses of the word 'logic'. Peirce's distinctions are a first step towards a clarification of what we mean by 'logic' when we say that logic is normative for thought.

4.4. On Peirce's use of the notions of *logica docens* and *utens*

In this section, I draw on Peirce's work on *logica docens* and *logica utens* to show how it could be applied to the debate on the relevance of logic to reasoning.

In his writings, Peirce revives a useful distinction within logic which first appeared in Scotus' discussion on syllogistics and then apparently had been forgotten for many centuries. The distinction at stake is between

Logica Docens, i.e. 'the result of scientific study' of arguments and of their constituent parts, the classification of arguments and, ultimately, the scientific theory of validity which can be learnt and taught; and

Logica Utens, i.e. a pre-scientific theory which enables us to reason according to some general principles and to recognize good and bad arguments. According to Peirce, we can't *properly* reason without employing such a logic.

Two considerations that come to my mind in analyzing these definitions are, first, that it may be questioned whether Peirce's definition of *logica docens* and *utens* fully matches with Scotus' and, second, how we ought to interpret the passage below in light of Peirce's discussion of the two senses of logic:

> I shall have a good deal to say about right reasoning; and in default of better I had reckoned that as a Topic of Vital Importance. But I do not know that the theory of reasoning is quite vitally important.

> That it is absolutely essential in metaphysics, I am as sure as I am of any truth of philosophy. But in the conduct of life, we have to distinguish everyday affairs and great crises. In the great decisions, I do not believe it is safe to trust to individual reason. In everyday business, reasoning is tolerably successful; but I am inclined to think that it is done as well without the aid of theory as with it. A Logica Utens, like the analytical mechanics resident in the billiard player's nerves, best fulfills familiar uses. (Peirce (1935): 108-109)

The above passage can be interpreted mainly in two ways, which differ substantially in the construal of the phrase '[T]heory of reasoning'. We will see how the interpretation of such expression has significant consequences in the right interpretation of Peirce's take on the connection between logic and rationality. According to Peirce's previous discussion of the topic, the options seem to be:

1. '[T]heory of reasoning' refers to *logica docens*. Therefore, the subsequent discussion is meant to argue that *logica docens* is dispensable in the 'conduct of life' and unsafe if followed by itself '[in] the great decisions', without rejecting that *logica utens* is necessary in our everyday reasoning. This interpretation of the passage would still be compatible with a view contrary to Harman's;

2. '[T]heory of reasoning' refers to *logica utens*. If so, Peirce would be alluding at the fact that normatively-guided inferential practice is dispensable in everyday reasoning, so his position would lean towards something analogous to Harman's view on the relation between logic and rationality.

In other passages[6], Peirce defines *logica utens* as a non-scientific inferential theory which every individual holds in performing reasoning, intended as the inferential activity of obtaining new knowledge from true premises, which makes *logica utens* look close to the mental logic we will discuss in the next chapter. In another passage, Peirce writes that 'Now a person cannot perform the least reasoning without some general ideal of good reasoning'.[7] Therefore, the construal of point 2 above seems implausible. In fact, as I mentioned, if a theory of good argument is a necessary condition for reasoning, reasoning 'without the aid of theory' would be enough for the reasoning to be irrational. In light of Peirce's discussion, I think that *logica utens* can be paraphrased as 'a pre-scientific theory of validity'. In fact, on Peirce's

[6]See Peirce (1932): 2.204.
[7]Peirce (1932): 2.186.

account *logica utens* is 1) a theory of what good reasoning should look like, therefore a normative theory; 2) a theory which nobody can dispense with when engaging in reasoning; 3) a non-scientifically informed theory. So, saying that *logica utens* is a folk (or pre-scientific) theory of validity doesn't seem too far-fetched.

However, it seems reasonable to ask what happens to an agent's *logica utens* when she comes to believe a scientific theory of validity (*logica docens*). Is *logica utens* superseded by *logica docens* once an agent has accepted a certain scientific account of validity?

If I am right in taking *logica utens* to be a folk theory of validity, I think that agents tend to retain *logica utens* anyway. But even so, what are the reasons for retaining a pre-scientific theory once one has accepted a scientific one? To answer this question, we might resort again to the passage above: 'In everyday business, reasoning is tolerably successful [.] [...] A Logica Utens, like the analytical mechanics resident in the billiard player's nerves, best fulfills familiar uses'.[8] Peirce seems to address the question, replying that the reason to keep a folk theory of consequence is that it is successful in our everyday reasoning. It might be so because *logica utens* is, e.g. inductively modelled and has a certain predictive power which our favourite deductive theory lacks. For instance, suppose I hold some truth-value-gap logic as my *logica docens* and that, according to that, I reject the Law of Excluded Middle (LEM), $\models A \vee\neg A$. Nevertheless, it might be wise to stick to LEM in certain ordinary cases, for example when I try to determine whether my cat has already been fed or not. As another example, suppose I hold classical logic, which entitles me to apply Ex Falso Quodlibet, $A \wedge\neg A \models B$, for any A and B. In my ordinary life, I don't really infer anything from a contradiction, but probably keep applying other logical laws which my favourite logic holds as valid.

Many have attempted to explain this seemingly double standard, such as the minimal rationality approaches discussed above. Those approaches defended a "human", situated model of rationality, in which principles of reasoning (which seem pretty close to epistemic norms) overcome logical principles. Here, I defend the view that *logica utens* offers a normative theory of good argument which ordinary reasoning must follow and that this view is compatible with (MR). As *logica utens* has normative powers and *logica docens* has too, they might be seen as alternatively overcoming each other. As Peirce writes, *logica utens* best fulfills familiar uses; on the other hand, it still might be that logically-trained agents resort to scientific theories of

[8] Peirce (1935): 109.

validity when engaging in reasoning about non-ordinary matters, more or less in the same fashion as people apply certain naïve physical theories when trying to predict roughly the path of a ball bouncing on the ground but have to resort to scientific physical theories when confronted with the path of objects in space.

In order to understand how *logica docens* and *logica utens* interact, it might be useful to consider in turn how each affects the other. Consequently, I will turn to examine first the case of *logica utens* affecting *logica docens* and then vice versa.

Case 1:

Our pre-scientific theory of good argument provides intuitions that we might want to transfer to *logica docens*. The motivation for developing some non-classical logics (and therefore, revising *logica docens*) has been that some classically valid inferences don't seem good inferences intuitively. A correlated question is: Can pre-scientific theories of good argument be revised? Intuitively, I would answer 'yes'. The *logica utens* an agent holds when she is, say, 17 years old, can change over time even without any training in formal or informal logic. Does revising *logica utens* affect *logica docens* somehow? I don't have a precise answer to this question, but in principle I think that it would be possible that revising *logica utens* affects *logica docens*.

Case 2:

Suppose that an agent gets training in logic for the first time, learns that some inferential patterns previously applied are invalid, that some seemingly counterintuitive inferential patterns are valid, etc. Then, the agent will presumably modify her *logica utens* accordingly, e.g. by eliminating from her *logica utens* the invalid inferences.[9] Nonetheless, as I mentioned above, agents might have reasons to retain their *logica utens* even after having learnt to apply a *logica docens*. Does revising *logica docens* affect *logica utens*? To answer, let's give an example. Suppose I have been trained to apply classical logic and to think that certain rules are valid, and suppose that helped me reshape my pre-scientific theory of validity. At some point, I realise that classical logic includes some invalid inferences and decide to adopt a paraconsistent logic. This might be motivated by my intuitions about the invalidity of Explosion, and in that case my decision about revising *logica docens* would perhaps be triggered by the logical intuitions in my *logica utens*. In cases like this, a revision in the *logica docens* doesn't usually affect much the *logica utens*.

[9]My intuition is that when agents learn to apply *logica docens*, they tend to reorganize their *logica utens* so as to eliminate invalid inferences more than to add counterintuitive valid inferences.

To conclude, let's consider models of rationality once again. The clash between (IAR) and (MR) has been largely motivated by the different take on logic - especially on the role and relevance of logic to rationality - they hold. (MR) focuses on actual agents, instead of ideal agents with infallible logical abilities. The whole minimal rationality project is concerned with developing a theory which takes into account the cognitive, memory and deductive limitations that real agents actually have.

On the other hand, as we have seen, (IAR) assumes that agents hold a certain logical system, which corresponds to what Peirce identified as *logica docens*, sticking to it without exception. Therefore, if such an ideal agent existed, her *logica utens* would conform exactly to the *logica docens* she holds. However, when minimal rationality approaches tackle the question about how relevant logic is to rationality, suddenly the options seem to squeeze into two alternatives: either we must expect from real agents to conform to *logica docens*, i.e. a scientific deductive theory of valid arguments, in all their beliefs, decisions, and intentions, or logic is not especially relevant to rationality.

As we have seen, from observing that real agents' reasoning doesn't conform to *logica docens* all the time, (MR) advocates conclude that it just doesn't conform to logic *tout court*. A possible response along the lines of Peirce's work could be that, unlike with ideal agents, real agents' *logica docens* and *logica utens* don't overlap entirely.[10] In fact, if *logica docens* is meant to be a scientific theory of validity, it is certainly the case that common people with no training in logic don't even hold a *logica docens*. This doesn't mean they don't have a clue about how good arguments ought to look like, though. Moreover, this doesn't mean that, for example, they lack the competence to deem as irrational something which is a *non sequitur*. So, what is this mysterious faculty which enables real agents to reason according to certain normative logical patterns? *Logica utens* could look like a good candidate for filling this gap, as it still displays the feature of being normative, which minimal rationality models want to preserve.

In the next chapter, we will address the question whether there is something like a *logica utens* applied by real agents. The corresponding psychological position is called Mental Logic theory, and holds that people reason in conformity to some mental inferential rules that, on some versions of Mental Logic, overlap with rules of classical logic.

[10]Despite I am applying the distinction to the discussion on minimal rationality, I should point out that the distinction between *logica docens* and *utens* is never mentioned by advocates of minimal rationality.

CHAPTER 5

Reasoning and conditionals

Abstract. Until late 1960's, the mainstream view in the psychology of reasoning held that human reasoning follows roughly the same principles as formal logic. The then-orthodox view is often called Mental Logic Theory. In 1966, Peter Wason developed his famous selection task to test people's deductive reasoning skills, and the results of his test became known as a knock-down counterexample to Mental Logic.

In the chapter, I distinguish the main reactions to the results of the test and I highlight that the responses to the selection task are philosophically relevant. Then, I focus on a particular strategy to rescue the core idea underlying Mental Logic, which I call the *Wrong Conditional Reply*, arguing that the strategy doesn't withstand scrutiny. To make my point, I consider alternative versions of the selection task (developed in the 1980's) with manipulated semantic content, and I stress that the responses given by the majority of people to those tests disconfirm the Wrong Conditional Reply.

In the last section, I discuss the main theories of reasoning and how they explain the evidence from experimental results, especially the selection task and its alternative versions.

Finally, I argue that if we agree that the selection task and its alternative versions are valuable as psychological evidence, we must conclude either that deductive reasoning is not indispensable for rationality, or that the responses given by people to the task show that they are often *arational*, rather than irrational.

5.1. Introduction

The last decades have seen a proliferation of studies in the psychology of reasoning. Until the half of the last century, scholars would generally agree with a strongly idealized view of human reasoning and, roughly speaking, with what philosophers could call a Fregean view of human reasoning competences.

However, some experiments performed in the late 1960's revealed a disconcerting deviance from logic in most of the subjects tested.

Since then, psychologists have tested people on a variety of reasoning abilities, e.g. deductive, conditional, and probabilistic reasoning, just to mention a few. Concomitantly, a wide array of theories have been developed to explain the experimental results.

This chapter will address some philosophically relevant conclusions that have been drawn from psychological experiments on reasoning. First, Wason's test will be described and it will be pointed out why the test is significant for the subsequent study of reasoning.

The selection task leaves many open questions for both philosophers and psychologists to answer. Since the results of the test have been published, the scientific community has been haunted by the suspicion that the test might prove that human reasoning is illogical and/or irrational.

A way to avert this conclusion is to contest that the selection task addresses the wrong kind of conditional, i.e. a conditional that is not often represented in human reasoning. I call this the *Wrong Conditional Reply*. If this response to the selection task were successful, it would be especially helpful to rescue Mental Logic theory, which is often deemed as the main victim of Wason's test.

According to the Wrong Conditional Reply, the responses to Wason's test could have been assessed taking into account conditionals different from material implication, consequently averting the conclusion that humans are not rational. Before discussing the Wrong Conditional view, Adams' and Stalnaker's probabilistic accounts of the conditionals will be summarized.

Then, the problems with some of the theories of conditionals mentioned will be pointed out.

In particular, Stalnaker's account of the conditional has come under strong criticism, and it has been subsequently ruled out by many. On the other hand, Adams' account of conditionals has had a better fate. Adams' conditional is the focus of a version of the Wrong Conditional Reply.

The Wrong Conditional thesis will be discussed and examined in light of the original version of the selection task and of the alternative versions of the task. Then, it will be determined whether non-material conditionals explain also the responses to the alternative versions of the selection tasks.

The last section will be devoted to the theories of reasoning that have been developed and spread, paying attention to the way those theories have respectively claimed Wason's experiment to be supporting evidence. But, before we start, it is worth pointing out what will be left aside from the discussion, as the experimental results we will consider in a moment are but a small part of all the studies produced in the psychology of reasoning.

There is a huge literature on reasoning, be it deductive or inductive, but not much work has been done to import the results from cognitive science to philosophy. The issue of human rationality raised from the psychological studies is important and philosophically relevant, but philosophers are sometimes discouraged from taking empirical results seriously.

However, there are some fortunate exceptions. The general issue of the philosophical import of reasoning experiments has been addressed by Stein (1997), who argues that a philosophical study of human rationality cannot dispense from taking into account the experimental results in the psychology of reasoning. In particular, Stein discusses what he calls 'the standard picture of rationality', and argues in favour of another view, that he calls 'the naturalized picture of rationality'.

Briefly, the standard picture of rationality is the view claiming that rationality is subject to some normative principles of reasoning, and the principles of reasoning are usually taken to be derived from logic or probability.

In other words, the standard picture is understood as taking logical or probabilistic principles as the standard against which reasoning is to be assessed. If a piece of reasoning doesn't match up with such normative principles, it can be judged as irrational. However, Stein argues, if we really want to keep the standard picture of rationality, we must accept the harsh verdict suggested by the experiments on reasoning.

In light of the experiments, Stein claims, there is no easy way to reconcile the standard picture with the view that human are in fact rational. Consequently, Stein goes on, arguing that humans are rational requires that the standard picture of rationality be ruled out.

For this reason, Stein suggests a different picture of rationality, which doesn't encompass a unique set of normative principles of reasoning. The naturalized picture of rationality is described as a system where the normative principles of reasoning are given by a wide reflective equilibrium, instead of being provided by logic alone (or probability alone). The reflective equilibrium keeps in balance what we consider correct reasoning; what we deem as the general principles of correct reasoning; philosophical and scientific theories, and scientific evidence.

Thanks to the scientific evidence, we acknowledge the cognitive and temporal limitations of human agents. In a naturalized picture of rationality, the scientific evidence input and the other inputs balance out. So, Stein claims, the picture of rationality emerging from the reflective equilibrium avoids the pitfalls both of a an idealized model of rationality, and of a trivial model of rationality.

5.2. Ideal agents and real agents

For some time, the study of reasoning in experimental psychology was dominated by the *Mental Logic* theory, according to which ordinary agents' reasoning conforms to inferential rules similar to those of classical logic. An early version of this view (also called *logicism* in the psychology of reasoning)[1] can be attributed to Piaget (1972b), who claimed that the adult cognitive stage is characterized by the ability to use symbols and abstract operations (the so-called *formal operation stage* of the cognitive development).

As we will see more in detail in the fourth section, the early developments of the psychology of reasoning were still clinging onto the view that reasoning was mostly syllogistic. It is more contentious whether the early studies referred to Aristotelian syllogistics, or rather to some modern re-elaboration of them.

Piaget was halfway between the early psychologists, who still identified human reasoning with syllogistic reasoning, and the modern psychologists. Therefore, it shouldn't surprise that Piaget's writings include plenty of references to syllogisms.

However, Piaget's work on operational logic shows that not only he was aware of the 20th-century developments of formal logic (for example, Piaget cites logical atomism and axiomatic theory), but he had substantial familiarity with modern logic.[2]

Piaget claimed that there is a correspondence between the formal operations of cognitively fully developed subjects on one side, and classical logic on the other. This is not the same as claiming that people were born with perfectly developed logical abilities, though.

[1] Not to be confused with *philosophical* logicism.

[2] For example, Piaget (1972a), §§49, 50 are devoted to discuss negation-free logics and multivalued logic, respectively.

With his famous *Cognitive Stages Theory* (CST), Piaget meant to show that in the last stage of cognitive development (that starts approximately at the age of 13), subjects are able to pursue and complete reasoning tasks requiring the ability to abstract away from concrete objects and situations. The last stage is called the Formal Operational Stage right because individuals at this stage display formal reasoning skills. In other words, subjects who have reached the last stage are able to reason on the basis of the *form* of propositions, i.e. the internal structure of propositions, regardless of their content. Morever, when a subject reaches the Formal Operational Stage, she is able to handle reasoning also with the relations *between* propositions, abstracting from the propositions' content.[3] CST was subsequently identified as an early variety of the broader Mental Logic theory and Piaget was discussed in the psychology of reasoning for his "Fregean" views on human reasoning.

The key points of Mental Logic theory had been taken for granted during the first half of the twentieth century at least but, from the late 1960's, experimental results started to cast doubt on the tenets of Mental Logic. The revolutionary experiments presented in Wason (1966) are taken to show that Mental Logic didn't get the picture quite right.

In the tests run by Wason, known as the *selection task*, the experimenter lays down four cards in front of you, as represented in *Fig.1*. You know that each card has a number on one side and a letter on the other side, and the experimenter asks you to turn over only those cards that are relevant to determine whether the following claim is true:

If a card has a vowel on one side, then it has an even number on the other side.

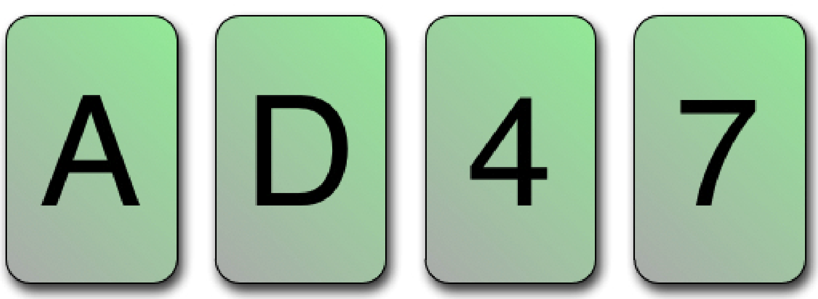

Fig.1[4]

[3]Piaget distinguishes *intrapropositional* operations from *interpropositional* operations. The latter are about relations between propositions (e.g. logical connectives), whereas the former refer to the internal structure of a proposition.

[4]Image from http://www.psychologyinaction.org/2012/10/07/classic-psychology-experiments-wason-selection-task-part-i/

It turns out that the vast majority of people taking part in the study select card *A* (which is correct) and a significant portion of them also pick card *4* (which is incorrect according to classical logic and most non-classical logics), whereas less than 10% of the participants turn cards *A* and *7* (arguably the logically correct choice).

The error made by people choosing to turn cards *A* and *4* in the test is seemingly to commit the *Affirming the Consequent* fallacy, whereas people who decide to turn only card *A* seem to be not capable of applying Modus Tollens. Therefore, the selection task has been revolutionary because it has been taken to show primarily that ordinary agents fail to reason according to logical rules substantively and repeatedly. However, not everyone agrees on what the psychological or philosophical import of the selection task is, and there have been many conflicting replies to the experiment.

To make things clear, we could divide the main reactions to the selection task into two big groups: in the first group, there are all those who think that there is a substantive way in which reasoning logically is a necessary feature of rationality (I call this view *Rationality&Logic*); in the second group, there are all those who think that following deductive reasoning is not an essential feature of rationality (I call this view *RationalityOrLogic*).

Within the Rationality&Logic group, there are two main responses to Wason's eperimental results. The first is just to take the results at face value and bite the bullet, admitting that human beings are illogical and, as such, humans are also irrational (we could call this the *irrationality thesis*, following Stein (1997)). I discuss this view in more detail in the fourth section.

The second response under the Rationality&Logic group tries to cope with the results without admitting that humans are irrational. There are two main strategies to do this, namely (i) arguing that the selection task is unreliable and the results of the test shouldn't be taken seriously, or (ii) arguing that the interpretation of the task incorrectly assumes that the conditional in the test should be interpreted as a material implication (I call this the *Wrong Conditional Reply*). The latter strategy supports the view that Wason's results are compatible with non-classical conditionals, such as, e.g. conditional probability or Adams' probabilistic conditional (see Duca (2009)).

Before we discuss the non-classical conditional response to the selection task, it is worth explaining what the conditional probability approach and, in particular, Adams' conditional are.

5.3. Conditional probability

Many of us are familiar with material implication, which has been taught in basic logic courses for decades. The truth conditions for material implication, $\supset$, are $v(\varphi \supset \psi) = 1$ iff $v(\varphi) = 0$ or $v(\psi) = 1$. The truth conditions say that the antecedent being false or the consequent being true are sufficient for the implication to be true, which might be found unintuitive. But perhaps the truth table for $\supset$ was enough to persuade us that we could ignore the seeming counterintuitivity of material implication and be content with it. However, even if we are convinced that material implication is an adequate account of the conditional, there are some examples that demonstrate that conditionals are not truth-functional. To see this point, I will use an example from Read (1995).

First of all, suppose you and I are discussing an issue and we have opposite views. As our views are mutually inconsistent, they cannot both be true. Now, consider the sentence

> Either if I was right so were you, or if you were right, so was I. (Read (1995): 73)

The truth tables for material implication and for disjunction tell us that one of the two conditionals must be true. Therefore, the disjunction is true, no matter who is right. However, we cannot help thinking that the conditionals are not true, so the example seems to give substance to the suspicion that conditionals are not truth-functional, after all.

In addition, Ramsey's test confirmed the intuitive worries about conditionals. The basic idea underlying Ramsey's test was: to prove the credibility of a conditional 'If φ then ψ', when you are uncertain about the truth of φ, add hypothetically φ to your set of beliefs. Then, check whether you would believe ψ. Clearly, the test is not designed to test the truth value of conditionals, but their chances to be believed.

Unsurprisingly, once put to the test, none of the conditionals in the example 'Either if I was right so were you, or if you were right, so was I' turns out to be credible. Consider the setup of our example once again: you and I have mutually inconsistent views on an issue, and each of us is arguing in favour of her view. Now, let's take the first conditional, 'If I was right so were you', and add hypothetically the antecedent ('I am right') to our stock of beliefs. Would you believe the consequent ('You are right')? Clearly not. Let's do the same with the other conditional, 'If you were right, so was I'. As Ramsey's test prescribes, let's take the antecedent ('You are right') and add it hypothetically to our set of beliefs. Would you believe the consequent ('I am right')? No, you wouldn't believe this either.

In response to the seeming implausibility of material implication, Stalnaker and Adams developed their conditional probability approaches which, as the name itself suggests, employ probabilities instead of truth values. The general considerations behind the choice of adopting probability are (a) that probabilities capture reasoning with uncertain premises better than truth conditions, and (b) that a theory of conditionals cashed out in terms of probabilities would recapture logical truths and falsities as absolutely credible or absolutely incredible propositions anyway.

Probability functions assign values between 0 and 1 to the propositions. Unsurprisingly, logical truths take value 1 and contradictions take value 0. Moreover, supposing that φ is a proposition, the probability of $\neg\varphi$, $p(\neg\varphi)$, is equal to the probability of 1 - $p(\varphi)$.

The conditional probability of ψ given φ (formally translated as $p(\psi|\varphi)$) is explained as the ratio of probability that both φ and ψ are true, divided by the probability that ψ is true, $\frac{p(\varphi\wedge\psi)}{p(\psi)}$, and provided that $p(\psi)$ is greater than 0.[5]

The conditional probability approach developed by Stalnaker (1970) suggests that conditionals be read as conditional probabilities. In other words, on Stalnaker's account, 'If φ then ψ' would be read as 'ψ, given φ'. Accordingly, $p(\text{if } \varphi \text{ then } \psi)$ is equal to $p(\psi|\varphi)$. Put in a slogan that has become familiar, the probability of the conditional is measured by the conditional probability.

Lewis (1973) opposes Stalnaker's account of conditional probability, proving that there is no proposition such that its probability is equal to the conditional probability.[6] In particular, Lewis argues, there is no conditional whose probability is measured by the conditional probability.

Adams (1975) offers a slightly different account of conditionals, that is aimed to avoid Lewis' objection to Stalnaker's account. In fact, Adams argues that the conditional probability measures the *assertability* of the conditional, rather than its probability.[7]

[5]See Read (1995) for a detailed discussion of theories of conditional.
[6]Lewis' argument against Stalnaker's thesis is often referred to as the *triviality argument*.
[7]See also Edgington (2008) for a nice summary of Adams' theory of conditionals.

In addition, Adams shows that an argument, $\Gamma \therefore A$, is valid iff the improbability of A[8] is smaller than or equal to the sum of the improbabilities of the propositions in Γ. In Adams' words:

> Given a probability function p for a factual language $\mathcal{L}$, and a formula φ of $\mathcal{L}$, the uncertainty of φ relative to p is the number $u_p(\varphi)$ = 1 - $p(\varphi)$. The uncertainty of φ measures the degree to which φ is regarded as unlikely. Two easily demonstrable facts about uncertainty as here characterized are that if φ entails ψ then $u_p(\psi)$ is no greater than $u_p(\varphi)$, and that $u_p(\varphi_1$ & ... & $\varphi_n)$ is no greater than the sum of the uncertainties $u_p(\varphi_i)$, for i = 1, ..., n. Combined, the foregoing imply that if a factual formula is a logical consequence of a finite set of such formulas, then the uncertainty of the consequence is no greater than the sum of the uncertainties of the premises. (Adams (1975): 49)

According to Adams' definition of validity, classically valid arguments are valid also probabilistically, provided that they have a non-conditional conclusion. Most notably, some classically valid inferential rules are valid in Adams' logic, such as φ, $\varphi \rightarrow \psi \models \psi$ (Modus Ponens), $\varphi \rightarrow \psi$, $\neg\psi \models \neg\varphi$ (Modus Tollens), and $\varphi \vee \psi$, $\varphi \rightarrow \chi$, $\psi \rightarrow \chi \models \chi$ ($\vee$-Elimination, or Reasoning by cases).[9] On the other hand, Adams' conditional doesn't validate many inference schemata having a conditional conclusion, such as $\psi \therefore \varphi \rightarrow \psi$, $\neg\varphi \therefore \varphi \rightarrow \psi$, and Conditional Proof.

With Conditional Proof invalid, some classical theorems can't be proven. For instance, $\varphi \rightarrow \psi \dashv\vdash \neg\psi \rightarrow \neg\varphi$ (Contraposition) is not a thorem in Adams' system. Reformulated in conditional probability terms, if Contraposition held, the probability of $\psi|\varphi$ would be equal to the probability of $\neg\varphi|\neg\psi$. However, $p(\psi|\varphi) \neq p(\neg\varphi|\neg\psi)$. Consider also the following sentence: 'If it should rain, it will not rain heavily'. The sentence doesn't intuitively imply its contrapositive, 'If it will rain heavily, it should not rain'.[10]

[8] The improbability of A is equal to 1 - $p(A)$.

[9] Here and in the remainder, I use $\rightarrow$ to denote Adams' conditional.

[10] It can be argued that the 'if' in the weather counterexample should be rather interpreted as an 'even if'.

It is important that we keep in mind these facts about probabilistic validity, because they are crucial to the discussion in the next section.

5.4. Conditionals in the selection task

It is now time to go back to the responses to the selection task. In the first section, we had seen that there are mainly two ways to react to the selection task's results: either you admit that rationality and logic sometimes go separate ways (I dubbed this option *RationalityOrLogic*), or you think that logic is an indispensable component of rationality (*Rationality&Logic*, in my slang).

Within the Rationality&Logic group, either you bite the bullet and admit that humans are irrational, or you hold tight to the idea that humans are rational. If you go for the latter, you have two options: you can choose to show that the selection task is not a reliable piece of evidence, or you can argue that the conditional in the task is erroneously identified as material implication. I have called the latter strategy the *Wrong Conditional Reply*.

According to this strategy, it is wrong to assume that the conditional claim in the test is a material implication. First of all, one could argue, material implication is an inadequate account of the conditional and it has been proven to be wrong (Stalnaker, Adams, and many others thought so, for example). Clearly, if you are convinced that conditionals are not truth-functional and you presuppose that humans do reason according to material implication, which is an inadequate account of the conditional by your lights, you are implying that human reasoning is rather hopeless. But, if the results of the selection task show anything, they show that the tested subjects don't apply material implication come what may. Therefore, your concern about people applying the wrong conditional, i.e. material implication, is ungrounded. In addition, if you are a non-truth-functionalist about conditionals, you cannot be assuming that humans *should* reason according to material implication, otherwise you would expect that humans conform to a wrong standard of reasoning.

On the other hand, one could suppose that the results of the selection task actually show that the tested subjects didn't read the conditional in the task as a material implication, but rather as a conditional probability. If it were so, the subjects would have intepreted the conditional claim in the task as something similar to 'Determine p(even number on one side/vowel on other side)'. This interpretation of the conditional claim would justify part of the responses to the selection task. To see this, suppose you want to determine the probability of a card having an even number on one side given that it has a vowel on the other side. Without a doubt, you choose the card having a vowel on its visible side, i.e. *A*, in our example.

However, the conditional probabilistic interpretation apparently doesn't give you a reason to turn the card with an odd number on it, as Contraposition fails with the probabilistic conditional. Therefore, conditional probability could explain why the tested subjects didn't choose to turn the *7*, along with the *A*.

Nonetheless, it does matter what account of conditional probability the Wrong Conditional Reply is ready to support. Some accounts of conditional probability, such as Stalnaker's, are subject to important criticism. In particular, Stalnaker's view has been disproved by Lewis' triviality argument. Consequently, granted that Stalnaker's conditional probability account is wrong, assuming that humans reason according to conditional probability *à la* Stalnaker would take one straight to the conclusion that human reasoning is structurally wrong. In such a scenario, the conditional probability reading of the conditional claim in the selection task wouldn't be in a better position than the material implication reading.

Clearly, a sensible choice of the conditional is crucial for the Wrong Conditional Reply to be effective. According to Duca (2009), the upshot of the selection task is indeed to show that human reasoning doesn't conform to material implication. Duca argues that Wason's results show that humans read the conditional claim in the test as an Adamsonian conditional probability.

In fact, the rule that the subjects tested fail to apply in the four cards task is Contraposition, i.e. $\varphi \to \psi \dashv\vdash \neg\psi \to \neg\varphi$. As we have seen, Contraposition holds with material implication, but it is not valid according to Adams' definition of probabilistic validity. Therefore, the subjects tested in Wason's experiment cannot be blamed for not applying a rule that is not valid in the theory of conditionals they use.

Duca adds that, unlike Contraposition, Modus Tollens (classically and probabilistically valid) is not among the rules that the tested subjects fail to apply. The reason for this is simple. The four cards task asks to test (therefore, to disprove) a conditional claim. As we know, Modus Tollens says that $\varphi \to \psi$, $\neg\psi \models \neg\varphi$.

However, in the selection task, the tested subjects are provided only with the conditional premise of the rule (i.e. $\varphi \to \psi$). As the second premise of Modus Tollens, i.e, $\neg\psi$, is missing from the input, the subjects are not to blame for not applying the rule.

As there are many criticisms showing that material implication is not the best theory of conditionals, Duca concludes that the tested subjects are in fact rational to *not* apply Contraposition. The subjects read the conditional in the test as a probabilistic conditional, and their interpretation of the conditional is right. Therefore, the subjects don't commit any reasoning error.

This hardly seems to be the end of the story, though. In fact, there is a consideration to be made about any theory rejecting Contraposition, and another consideration which is specifically about the version of the Wrong Conditional Reply discussed. First, a theory that denies the validity of Contraposition should also explain why others have thought Contraposition to be valid. A theory rejecting Contraposition might even admit that some applications of Contraposition are right, but it should provide an explanation about why it could be sometimes right to apply Contraposition.

In addition, there is another problem, that affects specifically the Wrong Conditional Reply. The Wrong Conditional thesis needs the results of the test to be consistent across the different versions of the selection task. Therefore, showing that people don't apply Contraposition in the vowel-even number task because they read the conditional as a probabilistic conditional is not enough to confirm the thesis.

In a famous alternative version of the selection task due to Griggs and Cox (1982), people are given a test having the same underlying structure as the four-card task. However, instead of letters and number, the cards describe concrete situations, such as 'Drinks beer', 'Drinks soda', 'Is 21 years old', 'Is 16 years old', and the claim to test refers to a concrete situation as well, such as 'If someone drinks alcohol, they must be over 18 years old'. As it happens, most people responded to this version of the selection task applying Contraposition (according to the testers, this was the correct answer), and the result of Griggs & Cox's experiment has been taken to suggest that the subjects tested intepret the conditional claim of the selection task contextually to the way the selection task is presented.

What the Wrong Conditional thesis doesn't explain is why people did apply Contraposition in Griggs & Cox's version of the selection task. On the contrary, it seems that trying to prove that people apply consistently the same conditional across the versions of the test ends up in a deadlock. But let's follow Duca's suggestion a bit further, and let's suppose for a moment that we agree with Duca that the subjects interpreted the conditional claim as a probabilistic conditional in the vowel-even number version. Then, we should admit that people apply Contraposition anyway in the underage-drinking version of the selection task. Let's concede that Contraposition is not valid with material implication exclusively, so we don't need to conclude that people read the conditional claim as an instance of material implication in the underage-drinking version of the test. Nonetheless, the experimental results would point out that the tested subjects are not consistent with respect to the application of the rule.

Furthermore, we have seen that Duca argues that it would not be rational to apply Contraposition in the selection task (Duca considers the vowel-even number test specifically). Then, either Griggs & Cox's test can be shown to provide weak evidence (for example, having methodological flaws, or similar reasons), or the Wrong Conditional response is committed to the irrationality thesis anyway.

If avoiding Contraposition is rational, yet people apply Contraposition in the alternative versions of the test, either (i) people could be blamed for giving an irrational response to the task in the alternative versions, or (ii) the theory rejecting Contraposition acknowledges correct applications of Contraposition, and needs to provide an explanation of these exceptions.

What Griggs & Cox's results show at first sight is that the subjects tested don't seem to intepret the conditional consistently over the different selection tasks. Therefore, the Wrong Conditional Reply is not an effective strategy to defend the Rationality&Logic view, unless it provides a plausible way to explain the exceptions, namely, the applications of Contrapositions.

There is still another option to be discussed: the irrationality thesis claims that people are irrational because they fail to apply Contraposition in the original version of the task. Nonetheless, we have seen that people apply Contraposition in the underage-drinking version of the test, which Griggs & Cox consider the right answer. If we are to show that people are irrational even though they seem to act rationally on some occasions, such as the underage-drinking test, we must prove that acting rationally occasionally is not enough to define human agents as rational, after all.

The irrationality thesis is addressed more in detail in the next section, and some possible responses to it are examined.

5.5. Errors of reasoning and their explanation

Experimental psychology has shown that human agents display a significant deviance from deductive laws in their reasoning. Wason's selection task is probably among the most famous examples of how human reasoning can go wrong in deductive tasks, although it was originally designed to test Piaget's theory that people reason according to an innate mental logic (that becomes fully developed around the age of 14) that happens to be classical logic.

However, the experimental results seem to show quite the opposite. The selection task was taken to show that human agents often fail to reason deductively. The question that stems from such evidence is: Can an agent make mistakes systematically in performing deductive reasoning and still be considered a rational agent? Moreover, performance in reasoning tasks includes a mix of correct reasoning and errors. How to account for that? The mistakes may be due to processing limitations, inaccurate heuristics, or comprehension difficulties. Wason interpreted the results as hinting at something more systematic and substantial about reasoning errors, namely that human agents are mostly *irrational*.

Nonetheless, this might sound as too harsh a conclusion. A charitable reading of Wason's conclusion could go along the following lines: we should draw a distinction between *arationality* and *irrationality*, and we could think of the notions of *arational* and *irrational* as mirroring the meanings of *amoral* and *immoral* in ethics. I will explain the reader the meaning of the latter pair of words in order to draw the parallel with arationality and irrationality.

An agent is (i) *amoral* if she is indifferent towards moral principles; (ii) an agent is *immoral* if she intentionally violates moral principles. For example, a person suffering from cognitive disorders who doesn't show empathy toward other human beings cannot be considered immoral. Such a person is considered amoral because she is not intentionally violating moral principles or social conventions; in fact, she is just failing to realize other people's needs and what behaviour is appropriate to meet those needs. Another case where people show amorality rather than immorality is when they deliberately suspend their moral considerations. For instance, when a scientist pursues her research without considering whether it will have good or bad consequences, she acts amorally. On the contrary, when a scientist is aware that her research will have a negative impact on her community or on the environment, but she carries on nonetheless, she is acting immorally.

Likewise, we could think of an agent as (i) *arational* if she reasons in a way that doesn't comply with the principles of rationality; we could think of an agent as (ii) *irrational* if she violates rationality principles consciously. In common usage, *irrational* is often used in a way that resembles the meaning of *arational*, which can explain why Wason concluded that human agents are irrational. However, as violating the principles of rationality intentionally is presumably something that agents don't do often, attributing *irrationality* to the people tested would be inaccurate. It is much more plausible that the subjects engaged in the reasoning task fail to see what rule they are expected to apply or, in other words, that they act arationally rather than irrationally.

The arationality hypothesis relies on an assumption: that we espouse the material implication reading of the conditional claim in the selection task. In fact, if we didn't take the material reading to be the correct one, we wouldn't ask whether the subjects tested exhibit the "symptoms" of irrationality or not.

So far, we have considered some of the views that can be attributed to the stance I labelled *Rationality&Logic*. In the remainder of this section, I briefly describe some of the theories that could be grouped under the RationalityOrLogic label.

5.5.1 Rationality Or Logic

As we know, the selection test has raised many questions on how its results should be interpreted. Wason's test has been used as supporting evidence for a variety of theses, and the very fact that the test apparently justifies diametrically opposed theories has even cast doubts on the reliability of the selection task itself.

O'Brien (1995) divides the reactions to Wason's experiment results into two large categories: on one side, the mental-logic theorists deny that the outcome of the selection task has any particular relevance to the Mental Logic, and refuse to take the experimental results at face value; on the other side, the bias theorists use Wason's task to prove that our reasoning abilities are content-sensitive. For example, the bias theory argues that the outcome of the selection task is unsuccessful when the prompt provided to the agents is abstract, whereas the agents conform to deductive laws when the test is presented in a less abstract and more familiar frame.

On the other hand, Evans and Over (1997) argue for a *dual system* explanation of the selection task's results, according to which we should distinguish two types of rationality, labelled $Rationality_1$ and $Rationality_2$.

The duties of the two rationality systems are described below:

Rationality$_1$ Thinking, speaking, reasoning, making a decision, or acting in a way that is *generally* reliable and efficient for achieving one's goals. This type of rationality doesn't require much cognitive effort from the agent, and the actions pertaining to Rationality$_1$ could follow behavioural patterns that the agent is not fully aware of.

Rationality$_2$ Thinking, speaking, reasoning, making a decision, or acting when one has a principled way, often sanctioned by a normative theory, for what one does. Rationality$_2$ is more demanding for the agent as it requires a conscious cognitive effort.

In this framework, reasoning errors detected by deductive tasks, such as the selection task, are taken to be a failure of Rationality$_2$. As this view assumes that the two kinds of rationality are completely separate, according to Evans and Over (1997), failing to pick the correct options in a deductive task doesn't imply that human agents are not rational *tout court*, but that they are *not rational*$_2$.

Although I am not committed to the frame suggested by Evans and Over, it should be pointed out that whatever the correct definition of rationality may be, the irrationality thesis doesn't seem to be the correct response to the results of the selection task, even if we endorse the material reading of the conditional in the experiment.

In the remainder, I sketch a summary of the psychological theories of reasoning that have been presented in the literature. Every subsection on the main theories of reasoning addresses also how the theory discussed replies to the challenge provided by the selection task and its versions.

5.5.2 An initial distinction in the study of reasoning: psychometrics and experimental psychology

Study of high-level cognition has originally been divided into two groups: psychometrics and experimental psychology. Both of them were developed at the beginning of the twentieth century and they have maintained separate paths until recently, when their complementarity started to be appreciated.

Psychometrics consists in the measurement of mental traits and, more famously, intelligence. The study of psychometrics has tried to isolate factors that were broadly associated with intelligence. Among such factors were the so-called *factor I* and *factor D*, which stand for 'induction' and 'deduction', respectively. Unfortunately, the methodology applied in psychometrics showed its faults when it became clear that psychometric tests didn't show any significant evidence supporting the existence of factors I and D. In fact, the experiments weren't sampled broadly enough for the resulting factors to reappear in further testing. In addition, the factors yielded by the psychometric tests were found to correspond to processes that were highly dependent on individuals.

On the other hand, the study of reasoning in experimental psychology follows many streams. The main theories of reasoning are:

a. *Mental logic*: rational agents reason conforming to inferential rules similar to those of classical propositional logic. This view can be attributed to Piaget (1972b,a), who claimed that the adult cognitive status is characterised by the ability of using symbols and abstract operations (the so-called 'formal operation stage' of the cognitive development).

b. Rational agents' deductive abilities are *content-sensitive*, as argued in Cosmides (1989, 2005); reasoning in humans follows certain patterns determined by cognitive biases, as argued by **?**Kahneman (2011).

c. *Mental models*: rational agents reason according to mental representations of possible situations, as held by **?**Johnson-Laird (2006).

d. Rational agents compute probabilities, as argued by Oaksford and Chater (2007) and Oaksford et al. (2008).

As mentioned earlier, Wason's selection task has been advocated as supporting evidence for different theories. Specifically, bias theory, mental model theory and the probabilistic-mind theory all claim the selection task as evidence in their favour, respectively.

Now, I want to discuss a bit more in detail the four competing parties in the psychology of reasoning. I will start with bias theory, then I will discuss mental models, mental logic and, last, the probabilistic mind theory.

5.5.3 Bias theory

Bias theory holds that human agents follow cognitive biases (especially *confirmation biases*) in their reasoning. For example, real agents tend to favour information that confirm their previous hypotheses (as concluded by Wason (1968)).[11]

A different branch of the bias theory distinguishes different levels in human cognitive architecture. In the study of human reasoning faculties, two levels have been individuated: (i) an *algorithmic* level that is responsible for computational processes and information processing; (ii) an *intentional* level that is not concerned with computation but rather with the goal of computation. Stanovich (2008) argues that errors in deductive tasks are due to the triggering of a set of autonomous processes[12] at the algorithmic level that we might identify as characteristic of heuristics. The response suggested from the autonomous processes can be overridden, and this is in fact what happens in a minority of cases (less than 10% of the responses on the selection task), and when the tested subjects are told what went wrong with their responses.

The overriding of the autonomous process is carried out by the cognitive operation of *decoupling*. Decoupling of cognitive representations consists in representing a belief as separate from the state-of-the-world it is representing. This is, among other things, what makes decoupling essential for hypothetical reasoning. Overriding the autonomous-process-primed response results in complying with the normative stance that requires that agents stick to decoupling of cognitive representations (decoupling is considered an algorithmic process). This is true even though sometimes failing to decouple triggers autonomous algorithmic processes that happen to result in the correct response to a test. For instance, when people respond to the underage-drinking task applying Contraposition,[13] they might be doing so because they don't detach the hypothetical reasoning from the frame the task is presented within. People are used to complying with the laws of the state they live in or, at least, they know that they can be punished for not abiding by those laws, and the underage-drinking example includes prohibitions that are familiar to them. Therefore, they respond correctly to Griggs & Cox's version of the test.

However, Stanovich argues that there is more than the failure of decoupling to explain the high rate of reasoning errors in deduction tasks. Individuals have epistemic values

[11]Although Wason took this to imply that reasoning biases are a sign of the irrationality of human agents.

[12]The processes in question are characterized by their ability to bring about a quick, mandatory response that doesn't depend on input from high-level systems and doesn't load the central processing capacity.

[13]Which is the correct response, according to the experimenters.

and norms that are processed at the intentional cognitive level, and that are often arranged as thinking dispositions, i.e. dispositions to, e.g. privilege beliefs already possessed over new beliefs (confirmation biases), or rather to challenge beliefs already held in favour of new beliefs. These thinking dispositions can predict argument evaluation skills, and they can lead to different responses in individuals with an equal algorithmic capacity.

Stanovich argues also against the widespread view that reasoning errors are to be attributed to a divorce between competence and performance in human reasoning. The competence/performance advocates assume that errors in the performance are random, whereas Stanovich (2008) argues that the epistemic norms at the intentional level are responsible for errors in most of the cases. Is this like saying that human agents have a principled way to go wrong on deductive tasks? In short, yes.

Stanovich's theory seems to be confirmed by evidence in syllogistic reasoning experiments. The following test was run on college students and was constituted of some examples of syllogisms, with the tested subjects asked to tell whether the syllogisms in question were valid or not. One of the syllogisms presented was:

> All living things need water
> Roses need water
> Therefore, roses are living things.

Around 70% of the students deemed the syllogism as valid, as reported in *Stanovich (2008)*. However, when the terms of the syllogism were replaced as follows:

> All insects need oxygen
> Mice need oxygen
> Therefore, mice are insects,

a substantial number of the students who deemed the previous syllogism valid, considered the second syllogism invalid.[14] In line with Stanovich, the outcome of the test can be taken to show that most of the students who responded to the task were complying with a possible epistemic norm such as 'Stick to reality as much as you can', formulated at the intentional level. Or, according to a different intepretation of the

[14]Experimental data reported in Stanovich (2008).

syllogism task's results, individuals might be inclined to think that if an argument has true premises and a true conclusion, it must be a sound argument.

We have discussed dispositions to reason or act in certain ways because of some cognitive processes taking place at the intentional level. Now, let's turn to the more general study of biases, which is not concerned directly with intentions but recognizes that dispositions play an important role in reasoning. Generally speaking, biases have been the object of a systematic study since the 1970's. Especially after Wason's work on the selection task, the notion of confirmation (or *matching*) bias was developed to explain why people responded to the task in the ways we have already seen. In particular, Evans and Lynch (1973) suggested that people tend to choose to turn cards *A* and *4* because they tend to match the values named in the conditional claim '*If a card has a vowel on one side, then it has an even number on the other side*' when they try to verify the claim.

The most famous studies in bias theory have been produced by Kahneman and Tversky. Kahneman and Tversky (1974) especially analyze judgment under uncertainty and argue that the biases that people display when they reason about the probability of an event can be grouped under three main classes. The categories in question are dubbed *Representativeness*, *Availability*, and *Adjustment and Anchoring*.

Here is a brief description for each of the biases classes:

Representativeness

Kahneman & Tversky explain that when people are asked about how likely is that a certain event will occur, or how likely is that an object belongs to a class, they apply a heuristic that can be called *representativeness heuristic*. In this heuristic, people often make use of stereotypes, as in an example cited by Kahneman & Tversky:

> [...] consider an individual who has been described by a former neighbor as follows: "Steve is very shy and withdrawn, invariably helpful, but with little interest in people, or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail." How do people assess the probability that Steve is engaged in a particular occupation from a list of possibilities (for example, farmer, salesman, airline pilot, librarian, or physician)? How do people order these occupations from most to least likely? In the representativeness heuristic, the probability that Steve is a

> librarian, for example, is assessed by the degree to which he is representative of, or similar to, the stereotype of a librarian. (Kahneman (2011): 954)

Strikingly, the use of stereotypes overrides numerical facts. In our example, it is statistically more likely that Steve is a farmer (or a salesman) because there are more people employed as farmers (or salesmen) than people working as librarians.

- *Insensitivity to prior probability of outcomes.* People engaged in an expected probability task often neglect the base/rate data coming from previous occurrences. In other words, when subjects are provided with information that activates expectations about an event, E, the information triggering the expectations trumps the information about the frequency of E occurring in the past. However, people tend to take prior probabilities into consideration when they are not provided with other information that activate expectations about E.
- *Insensitivity to sample size.* When people are asked to give a judgment on a property appearing in a sample, they apply representativeness heuristic. Given certain information about the probability distribution of a parameter, such as the height of male adults, the subjects tested tend to esteem the probability distribution of the relevant parameter in exactly the same way, regardless of sample size. Kahneman & Tversky mention a test to exemplify the bias:

> Consider the following example:
> Imagine an urn filled with balls, of which 2/3 are of one color and 1/3 of another. One individual has drawn 5 balls from the urn, and found that 4 were red and 1 was white. Another individual has drawn 20 balls and found that 12 were red and 8 were white. Which of the two individuals should feel more confident that the urn contains 2/3 red balls and 1/3 white balls, rather than the opposite? What odds should each individual give?

> In this problem, the correct posterior odds are 8 to 1 for the 4:1 sample and 16 to 1 for the 12:8 sample, assuming equal prior probabilities. However, most people feel that the first sample provides much stronger evidence for the hypothesis that the urn is predominantly red, because the proportion of red balls is larger in the first than in the second sample. (Kahneman (2011): 964)

The answers given to the task reveals that people tend to ignore the statistical fact that a deviance from average is more likely to be found in a smaller sample than in a larger sample. In addition, other tests have highlighted that people's responses seem to be unaffected by clauses that make the importance of sample size explicit.

- *Misconception of chance.* The bias of misconceiving chance can be seen as a special case of insensitivity to sample size. People expect random sequences of events to display the same features, regardless of the length of the sequence. For example, people looking at a roulette wheel spinning, will expect black after a long sequence of red, or they will expect red after a long sequence of black.[15] In other words, people expect the essential features of an event are represented uniformly, both locally and globally. In our example, subjects know that the odds that the roulette stops on black are equal to the odds of it stopping on red, so they expect an equal distribution of red and black even in a short sequence of spins.
- *Insensitivity to predictability.* People engaged in predictions about, e.g. the performance of a person, or the future profit of a company, tend to adjust their judgment tune to pieces of information that are irrelevant to the prediction. In a test reported by Kahneman & Tversky:

> subjects were presented with several paragraphs, each describing the performance of a student teacher during a particular practice lesson. Some subjects were asked to evaluate the quality of the lesson described in the paragraph in percentile scores, relative to a specified population. Other subjects were asked to predict, also in percentile scores, the standing of each student teacher 5 years after the practice

[15]Example taken from Kahneman and Tversky (1974).

> lesson. The judgments made under the two conditions were identical. That is, the prediction of a remote criterion (success of a teacher after 5 years) was identical to the evaluation of the information on which the prediction was based (the quality of the practice lesson). The students who made these predictions were undoubtedly aware of the limited predictability of teaching competence on the basis of a single trial lesson 5 years earlier; nevertheless, their predictions were as extreme as their evaluations. (Kahneman (2011): 970)

- *Illusion of validity.* In predicting a future event, people tend to overrate the importance of the match between the input information and the prediction output. The example of Steve the librarian shows that the role of a good fit between the input information and the prediction outcome is valued as of primary importance by people.
- *Misconception of regression.* Regression toward the mean is a statistical phenomenon that if a variable displays extreme parameters on its first measurement, it will display a smaller deviance from the average on its second measurement. Although the phenomenon is well known, Kahneman & Tversky claim that people don't find it likely to occur in many contexts where regression does occur. Otherwise, if people realize that regression occurs, they tend to give wrong explanations for it. For example, people tend to give causal explanations of regression phenomena (that are purely statistical, so causation is not involved). An example reported by Kahneman & Tversky:

> In a discussion of flight training, experienced instructors noted that praise for an exceptionally smooth landing is typically followed by a poorer landing on the next try, while harsh criticism after a rough landing is usually followed by an improvement on the next try. The instructors concluded that verbal rewards are detrimental to learning, while verbal punishments are beneficial, contrary to accepted psychological doctrine. (Kahneman (2011): 975)

Availability

The biases of availability occur when people implicitly correlate the importance of the evidence provided by a sample, and the ease to retrieve the sample. In other words, the easier is to think of a sample, the more important the evidence the sample provides is deemed. Under the 'availability biases' group, four categories of biases can be distinguished.

- *Biases due to the retrievability of instances.* When people are asked to judge the size of a group, their judgment is often influenced by the ease to recall members of the group. For example, in an experiment people heard a list of famous men and women, and were asked to assess whether the list included more famous men or more famous women. Some of the lists would contain names of women that were relatively more famous than the men, and other lists would contain names of men that were relatively more famous than the men. The subjects tested erroneously assessed the gender ratio in the lists on the basis of the number of popular names they could recall. (Kahneman (2011))
- *Biases due to the effectiveness of a search set.* People find some methods for retrieving instances of classes easier to apply than others. Accordingly, people tend to think that the easier search sets contain more instances than the more difficult search sets. Below is reported an experiment that highlights how the bias affects the ability to judge the size of a class.

  > Suppose one samples a word (of three letters or more) at random from an English text. Is it more likely that the word starts with r or that r is the third letter? People approach this problem by recalling words that begin with r (road) and words that have r in the third position (car) and assess the relative frequency by the ease with which words of the two types come to mind. Because it is much easier to search for words by their first letter than by their third letter, most people judge words that begin with a given consonant to be more numerous than words in which the same consonant appears in the third position. They do so even for consonants, such as r or k, that are more frequent in the third position than in the first. (Kahneman (2011): 980)

- *Biases of imaginability.* When a subject needs to imagine (instead of recall) instances of a class, she tends to correlate the frequency of the class with the ease to imagine instances of it. Therefore, classes whose members are easier to construct are judged as more frequently instanced than classes whose members are harder to imagine.
- *Illusory correlation.* When asked to judge the frequency of co-occurring events, people tend to overestimate the frequency whenever the events are linked by an intuitive or cultural strong association. In an experiment, originally conducted by Chapman & Chapman, and reported in Kahneman and Tversky (1974), people were presented with drawings made by hypothetical mental patients and the diagnosis corresponding to the patients. The subjects tested tended to overestimate the frequency of events traditionally correlated somehow, such as the patient's suspiciousness and characteristic drawing of the eyes.

Adjustment and Anchoring

The phenomenon of *anchoring* occurs when people try to estimate a quantity. More in detail, agents start from an initial value - which is a readily available number (the *ancho*r) - and set an interval including the anchor to reach a plausible answer. Experiments show that people don't shift too far away from the original anchor, letting the anchor bias their prediction.

- *Insufficient adjustment.* Anchoring can happen when the initial value is suggested by the test prompt or when subjects base their judgment on incomplete computations. Sometimes the anchoring effect has an insufficient adjustment as its by-product, as in the following experiment.

  Two groups of high school students estimated, within 5 seconds, a numerical expression that was written on the blackboard. One group estimated the product
  $8\times7\times6\times5\times4\times3\times2\times1$
  while another group estimated the product
  $1\times2\times3\times4\times5\times6\times7\times8$

> To rapidly answer such questions, people may perform a few steps of computation and estimate the product by extrapolation or adjustment. Because adjustments are typically insufficient, this procedure should lead to underestimation. Furthermore, because the result of the first few steps of multiplication (performed from left to right) is higher in the descending sequence than in the ascending sequence, the former expression should be judged larger than the latter. Both predictions were confirmed. The median estimate for the ascending sequence was 512, while the median estimate for the descending sequence was 2,250. The correct answer is 40,320. (Kahneman (2011): 988)

- *Biases in the evaluation of conjunctive and disjunctive events.* People have been shown to judge erroneously the probability of conjunction or disjunction of events. The experiment dubbed 'Linda the bank teller' has been taken to show that bias theory explains errors in iductive tasks. The experiment was first presented by Kahneman and Tversky (1983). It goes roughly as follows: Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable? 1) Linda is a bank teller, or 2) Linda is a bank teller and is active in the feminist movement? Although the probability of option 1 is significantly higher than the probability of option 2, 90% of the subjects tested chose option 2. Generally speaking, people have been found to 'overestimate the probability of conjunctive events and to underestimate the probability of disjunctive events.' (Kahneman (2011): 990)
- *Anchoring in the assessment of subjective probability distributions.* When asked about their confidence about a certain value, people tend to deviate from what decision theorists call *proper calibration.*

> A judge is properly (or externally) calibrated in a set of problems if exactly Π% of the true values of the assessed quantities falls below his stated values of X. For example, the true values should fall below X_{01} for 1% of the quantities and above X_{99} for 1% of the quantities.

> Thus, the true values should fall in the confidence interval between X_{01} and X_{99} on 98% of the problems. (Kahneman (2011): 429)

However, people don't usually apply the rule of proper calibration. Studies have shown that both naïve and non-naïve subjects judge roughly 30% of the cases they are asked to assess as smaller than X_{01} or greater than X_{99}.

In other words, people tend to be overconfident about their probability judgments.

General considerations about bias theory

As we have seen, bias theory applied to high-level reasoning tries to determine, for example, how variations in the presentation of the arguments affect the individual ability to detect, or to generate, deductively correct conclusions.

Bias theory results are meant to show that reasoning is affected by biases in a systematic and substantial way. Biases are heuristic tools common to many subjects of the species *homo sapiens* and that could have developed over time, perhaps for evolutionary reasons. Heuristics can be seen as cognitive shortcuts involuntarily used to avoid the overload of our limited capacity of information storage and retrieval, memory and cognitive power.

Although the main concern of Kahneman and Tversky (1974) is probabilistic reasoning, as shown by the biases discussed, the general picture Kahneman & Tversky argue for can be extended to explain errors in deductive reasoning tests, such as the selection task. Kahneman & Tversky suggest a *dual process* (or *dual system*) account of reasoning.

System 1 guides our thought by default; it is automatic and quick, it relies on heuristics and it is scarcely accurate. System 1 is responsible for the biases in our intuitions.

On the other hand, System 2 is slow and accurate, but it requires cognitive effort. System 2 can correct the immediate responses given by System 1.

From a bias theoretic point of view, the majority responses to Wason's selection task can be seen as an instance of *matching bias*. The matching bias is a heuristic that favours lexical matching over formal reasoning. In other words, people engaged in a deductive reasoning task tend to deem as relevant cases whose lexical content matches the content of the rules they are asked to confirm. For example, in the original version of the selection task, people associate the claim they are asked to test, i.e. 'If a card has a vowel on one side, then it has an even number on the other side', with card *A* and card *4*. This could explain why a significant portion of people answer to the task indicating that cards *A* and *4* should be both turned over.

However, it has been argued (e.g. by Rips (1994)) that the bias theory's predictive power is unsatisfactory, because it explains only a relatively small portion of the cases. For instance, the predictions of the so-called *atmosphere effect*[16] account for only the 44% of the tested cases. However, it can be replied that cognitive biases still explain a significant portion of reasoning errors in inductive reasoning experiments, and that the error theory provided by bias theory can explain errors in deductive reasoning experiments as well, such as the selection task. The substantial difference in the responses given by people for the different versions of the selection task seems to support the hypothesis that the way the problem is formulated changes the way people read the conditional claim in the task, such as in the underage-drinking example.

5.5.4 Mental model theory

The famous thesis stated in Wason and Johnson-Laird (1972) is that no existing formal calculus correctly models human inferences, and no purely formal system would succeed in doing so. However, it is contentious whether reasoning processes in the human mind are more sensitive to the syntactic features of arguments, or to their semantic features. Mental logic supporters argue that in human reasoning the syntactic aspect is prominent over the semantic aspect. Notably, Johnson-Laird (1983) argues for the opposite view, i.e. that the semantic content trumps the syntactic aspect of arguments. Johnson-Laird's thesis would apparently be confirmed by various experiments on deductive reasoning, where the subjects are asked to engage in tasks involving conditionals and their performances vary significantly depending on the content of the arguments presented, such as in the syllogism examples in the previous subsection.

Mental model theory argues that people reason envisioning possibilities. In other words, mental models are representations of situations (that can be real or not) and they are usually constructed to predict events. In a reasoning task, the premises are represented as the starting possibilities in a representation (model). Subsequently,

[16] The atmosphere effect is described as the triggering effect that certain situations have on producing particular behaviours, even when the behaviours triggered are inappropriate or pointless, such as gesturing when talking over the telephone.

the agent draws a conclusion on the basis of the possibilites envisioned. If a counterexample of the conclusion can be imagined, starting from the same premises, the conclusion is rejected.

In addition, mental model theory holds a dual-process view of human reasoning. Byrne and Johnson-Laird (2009) explains the point about the dual-process account with the example of a conditional, 'If she played a game then she didn't play music'. Process 1 (the quick-response system) triggers envisaging the model corresponding to the clauses that are explicitly stated in the conditional. Therefore, the agent immediately represents the situation *played a game - didn't play music.*

It takes more time to get a response from Process 2 (the slow-reasoning system), that enables the agent to represent all the possibilites compatible with the conditional claim. Relying on working memory, Process 2 make explicit the possibilities *played a game - didn't play music, didn't play a game - played music*, and *didn't play a game - didn't play music.*

The bedrock of mental model theory is summarized in Johnson-Laird (2008) as follows:

(1) Mental models are the bulk of reasoning and they are representations of possibilities.
(2) Mental models are usually *iconic*, which means that their structure maps the structure of the possibility represented.
(3) People consider mental models one at a time, to not overload their working memory.
(4) Mental models can be combined by conjoining possibilities compatible with the different premises.
(5) Mental models are guided by the so-called principle of truth: mental models represent only true possibilities and omit false ones.
(6) The interpretation of connectives is strongly influenced by semantic and pragmatic content.
(7) Different reasoning strategies are developed on an individual basis.

As already mentioned, reasoning consists of envisioning possibilities, and the possibilities are represented iconically. So, for example, if you tell me that your car is parked in front of the entrance door, I imagine one of the many possible situations where there is a car roughly in front of the door, perhaps with our cars parked next to each other. As there are many possible combinations that could represent the situation described, I imagine the possibility that has traits common roughly to all the possible situations.

Envisioning a possibility on the basis of some premises can also take the agent to draw a conclusion that hadn't been stated, as in the example reported by Johnson-Laird (2008). If you tell me 'The cup is on the right of the saucer' and 'The spoon is on the left of the saucer', I will envisage the row of items spoon-saucer-cup (left to right) and I will draw the conclusion that the saucer is in the middle, without it having been stated in the premises. (Johnson-Laird (2008): 208) Therefore, mental models don't represent only what is explicitly stated in the premises.

On the other hand, it is much more demanding to envision multiple possibilites. The process of representing multiple models relies on working memory, which has limited capacity, and makes more difficult to represent conjunctions, disjunctions, or conditional claims than atomic sentences. We have seen that mental models don't reflect exclusively what is explicitly stated in the assertions. This is true also for representations of composite assertions, with 'or', 'and', and 'if'. Models representing disjunctions, conjunctions, and conditional claims represent situations that are possible given the assertions, as in the already cited example about the football game.

One thing to remember about mental models is that in a combination of possibilities like a disjunction, the individual clauses are represented only when they are true. For example, consider the sentence 'The broadcast isn't on the radio or it's on cable TV' in Johnson-Laird (2008): 209. The example yields three possible mental models, namely *not radio-cable TV*, *radio-cable TV*, *not radio-not cable TV*. However, as people represent only true clauses in their models, in the actual representations, the false clauses will likely be omitted, therefore people will tend to envision *not radio-cable TV*, *cable TV*, *not radio.* The principle that pushes one to omit representations of false clauses from one's mental models is called by Johnson-Laird *principle of truth.* The principle of truth plays a role in economizing working memory, which is easily overloaded. Unfortunately, envisioning only true possibilities brings about some illusory inferences in our reasoning, i.e. logically invalid inferences that look as valid in light of the mental models applied.

But what about the mental representation of conditional claims, such as the selection task?

Johnson-Laird explains the response of the subjects tested in Wason's selection task pointing out that it is the proof that mental logic doesn't capture hypothetical human reasoning nor human reasoning in general. According to Johnson-Laird, the subjects' choice cannot ultimately be ascribed to the way the problem is formulated in the experiment. They argue that in the selection task, the incapability of most of the subjects tested to give the correct answer about the cards that need to be turned over, i.e. arguably, *A* and *7*, can be explained in terms of mental models. Johnson-Laird argues that people represent the original version of the selection task as an abstract model, where the salient choices on the basis of the conditional claim given ('If a card has a vowel on one side, it has an even number on the other side') are cards *A* and *4*. In the four cards experiment, people give the wrong response because they follow the principle of truth, which guides agents to envisage only the true claims and to leave the false claims aside. Therefore, to verify the conditional claim in the task, the subjects envision a model where the conditional claim is true, i.e. where the card has an *A* on one side and a *4* on the other.

An alternative version of the selection task that was reported for instance by Johnson-Laird (2008) and originally described by Wason and Shapiro (1971), was the first attempt to reformulate the experiment manipulating its content to test whether it affected the reasoning performance of the tested subjects. There, the subjects were asked to verify the claim 'Every time I go to Manchester I travel by train'. The cards presented for the test had, respectively, 'Manchester', 'Leeds', 'train', 'car', on their visible sides. As in the original version of the selection task, the subjects were asked to turn over the cards necessary to test the claim. As people were found to be more inclined to choose 'Manchester' and 'car' than they were to turn over 'A' and '7' in the original version of the selection task, the results of the test with manipulated content seemed to show that reasoning is sensitive to content.

However, mental model theory needs to explain more carefully why people have given correct answers to the alternative version of the selection task. We have seen that according to mental model theory, such a case would be a corroboration of the thesis that people are sensitive to the semantic and pragmatic content of the claims they are reasoning about. Nonetheless, model theory is not clear on why the tested subjects stick to the truth principle when they perform the vowel/even number task, but they are more inclined to overrun the principle when they perform the Manchester/train task. If the principle of truth underlies the general representation of models, people should envision the case where the conditional claim is true in both of the tests. Therefore, they should pick the options 'Manchester' and 'train', rather than 'Manchester' and 'car'.

In response to the difference in people's answers to the different tests, the mental model theory stresses that the context a conditional claim refers to can significantly change the way people represent the corresponding possibilities. As mentioned by Byrne and Johnson-Laird (2009), conditional claims like 'If she played a game then she didn't play soccer' trigger responses depending also on the knowledge the agent has. In our example, the possibilities represented are only two, i.e. *played a game - didn't play soccer, and didn't play a game - didn't play soccer*. The agent doesn't envisage the model *didn't play a game - played soccer* if she knows that soccer is a game. From examples like the soccer game, mental model theory concludes that conditionals are not truth-functional, as the interpretation of conditional claims doesn't depend exclusively on the truth value of the clauses in the conditional claims. On the contrary, the interpretation of conditional claims is sensitive to extra-formal features, such as the meaning and the context the claims refer to, or the knowledge the agent has about the situation described by the claim.

When mental models are affected by the agent's knowledge, as we have seen in the example 'If she played a game then she didn't play soccer', the phenomenon is known as *pragmatic modulation*. As Johnson-Laird explains,

> Consider this inference, for example:
> If Pat entered the elevator then Viv got out one floor up.
> Pat entered on the second floor.
> Therefore, Viv got out on the third floor.
> You envisage a possibility in which Pat entered the elevator on the second floor, Viv was already in it, the two of them traveled up together to the next floor up, the third floor, and then Viv got out. You infer this sequence of events from your knowledge of how elevators work. In such cases, pragmatic modulation adds information about temporal and spatial relations between the events referred to in a conditional [...] (Johnson-Laird (2008): 214)

The effect of knowledge on mental models has been shown to speed up the reasoning process sometimes, and to hinder reasoning other times. For instance, when a reasoning task involves content known to the agent, sometimes the application of some inference rules, such as Modus Tollens, can be facilitated by the content of the premises, as in the example (reported in Johnson-Laird (2008))

> If Bill is in Rio de Janeiro then he is in Brazil.
> Bill is not in Brazil.

From the premises, the subjects tested easily concluded 'Bill is not in Rio de Janeiro', which is correctly yielded by an application of Modus Tollens to the premises.

On the other hand, the content of the premises hinders the application of Modus Tollens in other cases, as in the example

> If Bill is in Brazil then he is not in Rio de Janeiro.
> Bill is in Rio de Janeiro.

From the premises, people are not inclined to infer 'Bill is not in Brazil', despite its being a logically valid conclusion.

From the observations that mental models are affected by meaning, context and knowledge, Johnson-Laird concludes that the interpretation that people give of the connectives is not truth-functional. Moreover, mental model theory takes into account that different people have different reasoning strategies, where a reasoning strategy is defined as 'a systematic sequence of elementary steps that an individual follows in making an inference.' (Johnson-Laird (2008): 216)

In sum, what mental model theory tries to show is that people rely on representations of possibilities when they reason and that effortful processes, such as envisaging more than one model, are kept to a minimum. The principle of truth enables the cognitive system to reason saving working memory, but it is responsible for some reasoning errors, such as the response of the majority of people in the original selection task.

Both bias theory and mental model theory oppose mental logic theory, that was apparently disproved by Wason's experiment. I haven't discussed the mental logic theory in detail yet, and it is now time to address it.

5.5.5 Mental logic theory

I summarized the main tenet of mental logic in 5.2 and mentioned that the results of the selection task has been read as disconfirming the plausibility of the mental logic theory. However, mental logic hasn't faded out completely after Piaget's work, and some have tried to counter the evidence from Wason's selection task showing that mental logic still is a plausible theory of reasoning. In relatively recent work, Rips argues that human reasoning follows some basic rules that display many similarities with natural deduction rules.[17] Certainly the experimental results we have discussed have led psychologists to draw the conclusion that human reasoning doesn't follow a logical structure but, according to Rips (1994), psychologists have judged those results hastily. To Rips, psychologists have largely overlooked the variety of logics developed in the last decades (logics that take notions like knowledge, belief, temporal precedence and succession, etc., into account). Consequently, psychologists have too hastily deemed the subjects in the tests illogical whenever those subjects failed to give a response conforming to *classical* logic.

Whereas the crucial notion in mental model theory was the notion of mental model, the crucial notion in mental logic theory is that of *mental proof*. As Rips explains,

> I assume that when people confront a problem that calls for deduction they attempt to solve it by generating in working memory a set of sentences linking the premises or givens of the problem to the conclusion or solution. Each link in this network embodies an inference rule that the individual recognizes as intuitively sound and that provides some control on search. Taken together, this network of sentences then provides a bridge between the premises and the conclusion that explains why the conclusion follows. (Rips (2008): 191)

Mental logic theory's fundamental claim is that people, although they may be not always successful in performing a mental proof correctly, nonetheless always engage in attempts to accomplish mental proofs when reasoning deductively. The mental logic theory has been interpreted as an attempt to restore psychologism about logic from

[17] See Rips (1994, 2008).

the psychologist's side, while a philosophical attempt to restore psychologism about logic has been made by Hanna (2006), who has argued in favour of a mild (Kantian) psychologism about logic. Hanna argues that human agents have a *protological faculty* and his main thesis is constituted by two subtheses: i) logic is cognitively constructed by rational animals, and ii) rational human animals are logical animals.

In the same spirit, Rips argues that people are provided with a basic inference system that includes some rules. The inferential rules are activated in the system's working memory whenever the agent needs to evaluate an argument or to draw a conclusion from some premises. In the former case (assessing an argument), the inferential system will work backwards, starting from the conclusion and trying to construct a mental proof by applying the inferential rules to the premises. In the latter case (drawing a conclusion from some premises), the inferential system will work forward, applying the rules to the premises available.

But how could reasoning go wrong, then? Mental logic replies that deductive reasoning errors, such as those in performing the original selection task, show that people lack the mental rule corresponding to Modus Tollens, which would enable them to respond correctly to the test. However, this reply is not convincing for a number of reasons. Were it true that humans have certain mental rules, e.g. Modus Ponens, but not others, e.g. Modus Tollens, wouldn't explain why people give correct answers when they perform alternative versions of the selection task, such as the underage-drinking test.

Moreover, Rips argues that psychologists have hastily judged mental logic wrong in light of the experimental evidence overlooking the non-classical logics that have been developed in the last decades. However, Rips doesn't support his thesis with a comparison of the experimental results with the non-classical logics available, leaving it ultimately unconfirmed.

In the last decades, the mental logic supporters and the mental model advocates have been opposing parties in a lively debate. However, Stenning (2002) argues that the dispute between the mental model and the mental logic approaches can't be resolved on empirical grounds. According to Stenning, the reasoning processes modeled within one system can be emulated within each of the other systems. Therefore, Stenning argues, the two theories are formally equivalent with regard to the cases studied.

Not all the theories of reasoning agree that deductive logic is the normative standard for reasoning. One such exception is the probabilistic theory of the mind, which

I am now going to discuss. We will see how it can be considered the inductivist counterpart of the mental logic approach, to some extent.

5.5.6 Probabilistic theories of the mind

The probabilistic theory of the mind, as formulated by Oaksford and Chater (2007), holds that cognition needs to be understood in terms of probability theory. The reason for such a claim is simple: ordinary reasoning is usually carried out in a context of uncertainty and is concerned with practical action on the basis of partial or unconfirmed information, which doesn't make applying deductive logic a feasible option for human agents.

One of the immediate questions about the probability approach is: if we opt for a probability theory of the mind, should we understand the probabilistic approach in terms of frequency or in terms of subjective probabilities? From a frequentist point of view, probability calculus can only be applied where *frequencies* (i.e. repeated events) occur, so a frequentist account would be of no help with regard to the probability of unrepeated events. However, as real agents seem to have degrees of belief about unrepeated events, the subjective account of probabilities (i.e. Bayesianism) would be a more sensible choice.

Oaksford & Chater stress that the results of the selection task have erroneously been interpreted as showing that humans are irrational because the test evaluates the responses against classical logic. In fact, the probabilistic theory of mind argues that what other theories of reasoning (most notably, mental model theory and mental logic theory) deem as reasoning errors, are not real reasoning mistakes. The context common to our everyday reasoning is uncertain, and human reasoning needs to be suited to deal with uncertainty accordingly. Therefore, people are used to reasoning with inferences as though they were defeasible; consequently, human reasoning is non-monotonic, i.e. from $\Gamma \vdash A$ it doesn't follow that $\Gamma \cup \{B\} \vdash A$.

According to Oaksford & Chater, this explains the majority of answers given to the selection task. In fact, Oaksford & Chater argue, the responses given to the selection task are compatible with an interpretation of the conditional claim in Wason's task in terms of conditional probability.

How should experimental results, such as the selection task and the *Linda* test, be interpreted according to probabilism? If probabilism replaces deductive logic in a theory of reasoning, Oaksford & Chater still need to explain why people seem to get probabilistic reasoning wrong.

To the evidence provided by studies on reasoning biases, such as Kahneman & Tversky's, Oaksford & Chater reply that

> Our accounts of human reasoning are framed at the computational level, that is, they characterise what is being computed not how (although we do provide a process account in terms of probabilistic heuristics for syllogistic reasoning). That people's behaviour well approximates the norms provided by these models, which are thereby descriptively adequate, does not necessarily mean that people are doing complex probabilistic computations in their heads. (Oaksford et al. (2008): 385)

In sum, Oaksford & Chater argue that the selection task and similar tests show just that the wrong normative standard has been applied in the study of reasoning.

However, there is still a problem with the probabilistic theory of reasoning. In section 5.4, we discussed an attempt to explain the experimental results from the selection task in terms of conditional probability. There, we argued that such a response needs to be tested against a wider collection of evidence. For example, versions of the selection task with manipulated content, such as the underage-drinking test, have shown that people do apply Contraposition. Therefore, in the manipulated versions of the test, people don't appear to read the conditional claim in the test as a conditional probability.

The same worry arises with Oaksford & Chater's view of reasoning. The probabilistic mind theory explains a small number of versions of the selection task. As Oaksford & Chater stress, the theory aims at drawing a normative theory of reasoning, arguing that reasoning needs to be evaluated against Bayesianism, rather than against classical logic.

However, if changing the normative standard explains unproblematically the evidence from the original version of the selection task and takes the failure to apply contraposition as optimally rational, it also raises a problem. In fact, the responses to other versions of the selection task where people don't appear to reason probabilistically count as reasoning errors on the probabilistic account. Once such evidence is acknowledged, it becomes unclear why one should prefer Bayesianism over deductive (or even classical) logic as a normative standard for reasoning.

5.6. Criticisms of the selection task and conclusion

Two main criticisms can be moved against the layout of the selection task.

A criticism of the selection task could move from a methodological consideration about the selection task. It has been pointed out that the selection task relies on fallibilism and on the hypothetico-deductive model as a normative standard. However - the skeptic might stress - the hypothetico-deductive model is currently deemed as outdated in the philosophy of science, with the result that its use to assess the deductive abilities of human agents might be seen as inadequate. To this objection, one could reply that it is not obviously relevant. In fact, tested subjects have proved to give responses conforming to falsificationism on alternative versions of the selection tasks, such as the underage-drinking test reported in Griggs and Cox (1982).

Second, one can question whether a test that presupposes material implication as the bulk of correct responses overlooks the existence of non-classical logics and of non-material conditionals. This argument could be supported for example by modern mental logic theory (as it is presented for example by Rips (1994, 2008)), which stresses the importance of taking into account non-classical logics, and by Oaksford & Chater's probabilistic theory of mind, which holds that the conditional claim in the test should rationally be read probabilistically.

However, this criticism seems to miss the point in two important respects. First, the rule that the tested subjects fail to apply (i.e. Contraposition) is not an exclusive feature of material implication; on the contrary, contraposition is endorsed by many non-classical accounts of the conditional as well.

Second, evidence from alternative versions of the selection task has shown that people don't seem to read coherently the conditional claim across the different formulations of the test. Therefore, arguing that the normative standard for conditional reasoning should be non-contrapositive is not in a better position than arguing that conditional reasoning should be contrapositive.

Therefore, what we called the Wrong Conditional Reply to the selection task is not a successful way to rescue the view that reasoning logically is a necessary feature of rationality (we dubbed this view *Rationality&Logic*). If we agree that the selection task and its alternative versions are valuable as psychological evidence, we must conclude either that deductive reasoning is not indispensable for rationality, or that the responses given by people to the task show that they are often *arational*, rather than irrational.

CHAPTER 6

Logical norms and the epistemic paradoxes

Abstract. Some of the current attempts to account for the normativity of logic try to formulate epistemic norms in a conditional form, such as the principles discussed in MacFarlane's *In What Sense (If Any) Is Logic Normative For Thought?* (2004) and Restall's *Multiple Conclusions* (2005). The main *desideratum* for these principles is that they provide a plausible way to explain the connection between valid arguments and informal reasoning, i.e. reasoning that could be attributed to a rational agent.

In this chapter, I draw a distinction between logical laws and logical norms, then I show that none of the logical norms discussed is able to both (i) satisfy the *desideratum* and (ii) provide requirements for rational belief.

The norms I take into consideration are the normative requirements discussed by Broome in his article *Normative Requirements* (1999), the bridge principles presented by MacFarlane, and the principle discussed by Restall.

I make my assumptions explicit, then I proceed to considering the Preface and the Lottery Paradox and test the logical norms against them, showing that the norms formulated as requirements for belief don't provide rationally plausible principles. The logical norms formulated as *reasons* for belief fare better with the epistemic paradoxes than their strictly prescriptive counterparts. However, it can be argued that they don't have sufficient constraining power on our beliefs.

Therefore, none of the bridge principles is able to both (i) provide a plausible way to connect valid arguments and informal reasoning, and (ii) provide requirements for rational belief.

In the last section of the chapter, I argue that the kind of normativity logical norms belong to is much weaker than the normative requirements and the analogous principles would suggest.

What is the normative force of logical norms? Some of the current attempts to account for the normativity of logic try to formulate epistemic norms in a conditional form, such as the principles discussed in MacFarlane's *In What Sense (If Any) Is Logic Normative For Thought?* (2004) and Restall's *Multiple Conclusions* (2005).

The main *desideratum* for these principles is that they provide a plausible way to explain the connection between valid arguments and informal reasoning, i.e. reasoning that could be attributed to a rational agent.

In this chapter, I first present a classification of norms drawn from Glüer and Wikforss (2010), then I draw a distinction between logical laws and logical norms, and show that none of the logical norms discussed is able to both (i) satisfy the *desideratum* and (ii) provide requirements for rational belief. The norms I take into consideration are the normative requirements presented by Broome's *Normative Requirements* (1999), the bridge principles discussed by MacFarlane (2004), and the principle discussed by Restall 2005.

Most of the principles I consider share a common form, that is:

> If $\Gamma \models \Delta$, then you ought/are permitted/have reason to believe/not disbelieve [the disjunction of the propositions in Δ if you believe all the propositions in Γ].

I first consider an instance of the Preface Paradox and test the logical norms against it, showing that the logical norms formulated as requirements for belief don't provide rationally plausible principles.

It may seem at first that the Preface Paradox calls for either of the following options: (i) dismissing the Agglomeration Principle (i.e. $Bp \wedge Bq \rightarrow B\ (p \wedge q)$), or (ii) dismissing the logical norms in trouble, and the probabilistic solution to the paradox has been assumed to provide reason to reject the first option.

However, I subsequently discuss the Lottery Paradox and show how the probabilistic solution to the Lottery Paradox gives one ground to reject even a weaker version of Agglomeration, which I call *Weak Agglomeration* ($Bp \wedge Bq \rightarrow \neg\ Disbelieve(p \wedge q)$). I highlight that Weak Agglomeration is equivalent to the logical norms in trouble.

The logical norms formulated as *reasons* for belief fare better with the epistemic paradoxes than their strictly prescriptive counterparts. However, it can be argued that they don't have sufficient constraining power on our beliefs.

Therefore, none of the bridge principles is able to both (i) provide a plausible way to connect valid arguments and informal reasoning, and (ii) provide requirements for rational belief.

In the last section of the chapter, I consider the classification of norms again and argue that the kind of normativity logical norms exhibit is much weaker than the normative requirements and the analogous principles would suggest.

6.1. Logical laws and logical norms

When we talk about the laws of nature (e.g. thermodynamic principles, chemical bonding laws, etc.), we presume that an extended effort has been necessary to abstract those laws from observations of real phenomena.

For instance, the First Law of Thermodynamics - i.e. 'The increase in internal energy of a closed system is equal to the difference of the heat supplied to the system and the work done by it' - describes the correlation between internal energy, temperature and work of a physical system. If any measurements reported data conflicting with the First Law of Thermodynamics, and the scientific community agreed that the measurements are reliable, the data would cast substantial doubt on the law.

In other words, the laws of thermodynamics are not laws in the sense that they have the power to, say, prohibit that heat flow spontaneously from a colder location to a hotter location the same way as a law of a state has the power to prohibit cultivating marijuana. Like the other laws of nature, the laws of thermodynamics are the outcome of a long series of observations of natural phenomena and the result of a progressive dismissal of unconfirmed hypotheses about those phenomena.

The laws of nature spell out both the conditions for a phenomenon to occur and what a phenomenon is. For instance, the mere perception of a lightning doesn't tell us much about the nature of lightnings. It isn't until we learn that lightnings are electrostatic discharges cloud-to-cloud or cloud-to-ground that we know what is going on when we see a flash in a stormy sky.

On the other hand, the laws of a state (e.g. written rules about speed limits, laws about smoking in work places, etc.) are seldom concerned with explaining or clarifying the actions they regulate. Laws about what is legal and what is illegal have a constraining power on the citizens' behaviour. Unlike the First Law of Thermodynamics, the law that prohibits selling alcohol to underage people is not called into question if a boy of fifteen manages to buy some beer, and so violates the law.

In Normalese, laws of this type are sometimes called *norms for action*. They are usually norms that have an end independent of themselves; for example, legal constraints are enforced to guarantee that some rights be protected, or to ensure that people don't damage their own and others' health.

Frege explains the distinction between descriptive and normative laws in the Introduction to his *Grundgesetze* (Frege (2013) : xv), where he characterizes the former as laws describing the regularities in the world, and the latter as laws prescribing a course of action or providing a standard against which we judge what we do or what we think.[1]

Unlike laws with descriptive/explanatory power, such as the laws of nature, logical laws are typically taken to exert a normative pull on agents. However, a realist could oppose a characterization of logical laws as merely normative laws. In fact, if we agreed with her that the core logical notions, such as validity and consequence, are about *facts* and are mind-independent, the laws of logic would deserve to be acknowledged as descriptive laws, as well. This is exactly the position held by Frege, who argues that logical laws are in fact both normative and descriptive.

Although the issue of the status of logical laws is beyond the focus of this chapter, I want to clear the field from the ambiguities that could arise from the realist-antirealist dispute about the status of logical laws. To this extent, in the remainder I use *logical laws* to refer to the laws of propositional and predicate logic, such as the Law of Excluded Middle, with no regard to their normative role on thought; and I use *logical norms* to refer to the principles that connect deductively valid arguments to reasoning, giving a normative twist to logical laws.

What I mean by *logical norm* encompasses, e.g. the bridge principles discussed in MacFarlane (2004), but also other principles that are meant to provide necessary requirements for rational belief, such as the principle outlined in Restall (2005).

I hope to show that logical norms that try to map valid arguments to rational reasoning, such as the bridge principles presented in MacFarlane (2004), fail to provide necessary requirements for rationality. To this extent, I first present a taxonomy of norms, borrowed in part from the classification found in Glüer and Wikforss (2010), which I slightly modify to better account for the normativity of logic. Next, I apply instances of the Preface and the Lottery Paradoxes to put to the test the logical norms I discuss in section 6.3. In the last part of the chapter, I ask how logical

[1]See the discussion on Frege's account of laws in MacFarlane (2002).

norms fit into the taxonomy of norms, and I argue what kind of norms logical norms are.

6.2. Categories of norms

Extensive work has been done on norms, but we need to figure out what kind of normativity logical norms have. A general distinction between the manifold of norms will come in handy in this endeavour. In their entry on meaning and content normativity, Glüer and Wikforss (2010) distinguish several types of norms. Before we discuss the taxonomy of norms, it is worth pointing out what we will leave aside in our discussion, as Glüer and Wikforss (2010) themselves do in the entry.

The literature on normativity has seen two large parties opposing each other regarding the status of norms. On one side, cognitivists argue that norms are factual and mind-independent so, in other words, cognitivists are realists about norms. On the other side, many non-cognitivists argue that normative statements are not factual, therefore they don't have truth conditions.[2]

A classification of norms can be drawn independently of one's views about the metaphysical status of norms. Therefore, we won't need to get into the debate on the factuality of norms for the purposes of this chapter. The main distinction between norms that is reported in Glüer and Wikforss (2010) is between norms for action and norms of being.

As the names suggest, norms for action regard what should be done. In this sense, norms for action can be considered *directive* norms because they are supposed to guide or direct our actions. On the other hand, norms of being are norms that can be used to evaluate phenomena or behaviours. Norms of being are evaluative because they provide a standard against which a situation can be assessed.

This distinction between norms for action and norms of being resembles the distinction between the *first-person* and *third-person* use of norms mentioned by Pollock (1987). According to Pollock, a norm is used in a third-person way when it is applied to judge someone else's behaviour, skills or capacities. A norm is used in a first-person way when it is applied to guide and direct our own actions, as in sorting out what word we should use in a specific context.

In effect, the correspondence between the first-person use of norms and the norms for action on one hand, and between the third-person use of norms and the norms of

[2]However, this is not true of some non-cognitivists, like Field, Gibbard, and Blackburn, who think that normative statements have truth conditions or something closely resembling truth conditions.

being on the other, seems to be close to a case of synonymy, more than a resemblance. Therefore, in the remainder, we could harmlessly replace *norms for action* with *first-person norms,* and *norms of being* with *third-person norms.*

The classification presented in Glüer and Wikforss (2010) is specifically concerned with the norms for action, and it includes a large family of subcategories of those norms. In effect, norms for action can be subdivided into four main categories: let's begin by discussing the first two.

Norms for action can be instrumental or non-instrumental. Instrumental norms are directed toward accomplishing something and they depend upon the agent's goal, as in 'I should run everyday to be able to run the London marathon', or as in 'I should open the window to let some fresh air in'.

Non-instrumental norms, on the other hand, are not functional to make the agent attain a goal, as in 'You shouldn't talk during a play'. Non-instrumental norms can be constitutive rules or prescriptions. We will come back to this in a moment.

Before we discuss the subdivision of non-instrumental norms, we need to mention two more categories the norms for action are divided into. Besides being instrumental or not, norms for action can be *prima facie* or categorical. The difference between these two kinds of norms lies in their normative strength. Categorical norms are obligations that don't admit conflicting obligations. Therefore, categorical norms cannot be overridden by other norms. On the other hand, *prima facie* norms are compatible with other norms prescribing opposite actions. *Prima facie* norms can be overridden, given certain conditions, as in 'You shouldn't overtake a car on its left'[3] (not enforced when the car before yours is turning right), or as in 'You should keep secrets' (not enforced when a serial killer tells you about her crimes, for instance).

Let's go back to non-instrumental norms. We have seen that they can be prescriptive or constitutive norms. Let's discuss constitutive norms first. Constitutive norms (or rules) regulate activities that owe their existence to the norms themselves. For example, rugby rules are constitutive because the game of rugby wouldn't exist without them. An example of a constitutive rule is 'In the game of rugby, grounding the ball in the opposition ingoal area counts as scoring a try'. On the other hand, a prescriptive norm is a norm that could be formulated as an obligation, a prohibition, or a permission to do something. According to Glüer and Wikforss (2010), the main feature of prescriptions is that they can be expressed in deontic language.

[3]Assuming that you drive in the UK.

In turn, prescriptions can be conditional or unconditional. Conditional prescriptions are prescriptive norms that become enforced when a specified condition occurs, as in the already mentioned example 'You shouldn't talk during a play'. In the example, the condition that you are attending a play enforces the norm that you should keep quiet. Another example of conditional prescription is 'If someone holds the door for you, you should thank them'. Unconditional prescriptions, on the other hand, don't have any enforcement conditions. Some examples are 'You should comply with the laws', or 'You should donate to charities'.

With the categories of normativity in mind, let's turn to discuss the logical norms that will be examined and put to the test later.

6.3. Logical norms...

Some relatively recent accounts of the normativity of logic, such as Broome (1999), MacFarlane (2004), and Restall (2005) attempt to formulate logical norms in a conditional fashion. The main *desideratum* for such logical norms is that they provide a plausible way of explaining the connection between deductively valid arguments, and informal reasoning, i.e. reasoning that could be attributed to a rational agent.[4]

Most of the logical norms (bridge principles and similar norms) share a common form:

> If $\Gamma \models \Delta$, then you ought/are permitted/have reason to [believe/not disbelieve the disjunction of the propositions in Δ if you believe all the propositions in Γ].

In the following, I will consider Broome's normative requirements, MacFarlane's bridge principles, and what I call the *Consequence Principle*, and I will argue that these logical norms either (i) don't meet the *desideratum*, or (ii) don't provide requirements for belief.[5]

Let's start with Broome's normative requirements.

6.3.1 Normative requirements

[4] I stick with the meaning of 'informal reasoning' discussed by Harman (1986). Harman identifies reasoning with reasoned change in view, which, he argues, cannot be attained by applying logic alone.

[5] Not to be confused with recommendations for belief, as we will see later.

The basis for the views on the normativity of logic we discuss in this chapter have been laid by Broome, who has argued that normative requirements need to be distinguished from reasons. According to Broome (1999), the fundamental difference between 'having a normative requirement to φ' and 'having a reason to φ' is that a reason to φ is *pro tanto*. This means that having a reason to φ is compatible with having a reason to not-φ. More recently, Broome has defined *pro tanto* reasons as 'something that plays the for-F [where F is an action] role in a weighing explanation of why N [where N is an agent] ought to F, or in a weighing explanation of why N ought not to F, or in a weighing explanation of why it is not the case that N ought to F and not the case that N ought not to F.' (Broome (2013): 53)[6]

On the other hand, having a normative requirement to φ implies that you have an *obligation* to φ. Broome argues that normative requirements are best represented as wide-scope obligations. For example, if you have a normative requirement over a conditional statement, $\varphi \to \psi$, the normative requirement you have shouldn't be expressed as $\varphi \to O\psi$, but it should rather be formalized as $O(\varphi \to \psi)$ (which could be informally read as 'You ought to see to it that if φ, then ψ').

According to Broome (1999), logical consequence imposes normative requirements on rational beliefs, i.e. an agent has a normative requirement to believe the conclusion of an argument if she believes its premises. In other words, whenever it is obvious that $\Gamma \models A$, an agent S has a normative requirement to believe A provided that she believes all the propositions in Γ.

As we have seen in the third chapter, there is a quick argument explaining why Broom thinks that it would be a mistake to read normative requirements as narrow-scope obligations, i.e. $\Gamma \to OA$. Consider $A \models A$, which is a theorem of (virtually) any logic. If the normative requirements were narrow-scoped, we would have that 'If you believe A, you ought to believe A'. However, one could just be wrong in believing A, so the mere fact of believing A shouldn't suffice to enforce an obligation to believe A.

Therefore, the correct formalization of the requirement is wide-scoped, i.e. $O(\Gamma \to A)$, that can be informally read as 'You ought to see to it that if you believe all the propositions in Γ, you believe A'. Broome's argument in favour of wide-scope

[6]Although more recent discussion on normativity distinguishes *pro tanto* reasons and *pro toto* reasons, where the latter are defined by Broome (2013) as reasons explaining why you ought to do something.

requirements hasn't gone unchallenged [7] but, for the purpose of this chapter, I will leave the debate about wide-scope and narrow-scope requirements aside. The ultimate goal of this chapter is to test some logical norms that have been formulated explicitly in the literature and test their plausibility.

Besides the normative requirements and the Consequence Principle (presented in 6.3.3), I also discuss the bridge principles introduced by MacFarlane (2004), which are the object of the next subsection. For the purpose of this chapter, I focus especially on the principles MacFarlane argues for, leaving the remaining principles aside.

6.3.2 MacFarlane on bridge principles

MacFarlane (2004) distinguishes eighteen bridge principles. The bridge principles are classified on the basis of their

(i) *deontic operator* (ought to/permitted to/has reason to);

(ii) *scope* of the deontic operator;

(iii) *affirmative/negative form*.

The general form of the bridge principles is

> If A, $B \models C$, then [normative claim about believing A, B, and C - nested conditional].

The three main families of bridge principles are classified on the basis of their deontic operator's scope. Below, the scope of the deontic operators in the different families of principles is coloured in red.

- C principles:[8] If A, $B \models C$, then *if you believe A and you believe B, you ought/are permitted/have reason to believe/not disbelieve C*.

[7] For example, Kolodny (2005) argues that it is not straightforward that rationality requirements should take the wide scope and shows how the arguments for wide-scope requirements are ultimately unsuccessful.

[8] 'C' stands for 'consequent' of the nested conditional.

- B principles:[9] If A, $B \models C$, then *if you ought/are permitted/have reason to believe A and to believe B, you ought/are permitted/have reason to believe/not disbelieve C*.
- W principles:[10] If A, $B \models C$, then *you ought/are permitted/have reason to see to it that if you believe A and you believe B, you believe/not disbelieve C*.

Every family includes six principles, i.e. an affirmative and a negative principle for each deontic operator.

Before we go on with our discussion of the logical norms, it will be worth spending a few words explaining what the difference is between *not-believing that* p and *disbelieving that* p. We have seen that the possible propositional attitudes concerning belief are believing, not-believing, and disbelieving. Now, let p be the proposition 'Lincoln was shot in 1865'.

If we say that an agent doesn't believe that p, we mean that the agent *fails* to believe that Lincoln was shot in 1865. However, the agent not believing that p doesn't imply that the agent believes that Lincoln was *not* shot in 1865, for example because she might just feel unsure about the year of Lincoln's death. In other words, the agent might be lacking the belief that p as well as the belief that *not-p*, or she might be (voluntarily) withholding belief that p. Therefore, the negation isn't imported from the agent's attitude (*not* believing that p) to the proposition itself (believing that *not-p*).

On the other hand, when we say that an agent *disbelieves* p, we mean that the agent believes that p is false or, in other words, that the agent believes not-p. In our example, if the agent disbelieves p, she believes that Lincoln wasn't shot in 1865.[11]

Now, it is crucial that we understand what happens when we negate these propositional attitudes. Let's start with *not-believing* p: saying that it is *not true* that the agent *doesn't believe* that p just means that the agent believes p. Double negation makes it the case that *not-not*-belief that p is equivalent to belief that p.

[9] 'B' stands for 'both' the antecedent and the consequent of the nested conditional.
[10] 'W' stands for 'whole' nested conditional.
[11] It is worth pointing out that there is a well known pragmatic-linguistic phenomenon, the so-called *Neg-Raising* phenomenon, that affects the use of negations, as when we express disbelief about p by saying that we don't believe that p.

The story is different for *disbelieving* p. If we say that an agent doesn't disbelieve p, it means that the agent doesn't *reject* p, which is not equivalent to believing (or accepting p). Not-disbelieving p rules out only the possibility of rejecting p, but it doesn't prevent one from suspending one's judgment about p.

That said, we can now go back to the bridge principles, where the notions of disbelieving and not-disbelieving are employed frequently.

As we have seen in the third chapter, MacFarlane argues that most of the eighteen principles don't withstand scrutiny. For the purpose of this chapter, I don't examine MacFarlane's argument again nor survey all the principles.[12] I limit myself to remind the reader that MacFarlane takes his argument to show that the following three principles are the best options:

Wo- If A, $B \models C$, then you ought to see to it that if you believe A and believe B, you don't disbelieve C;

Wr+ If A, $B \models C$, then you have reason to see to it that if you believe A and believe B, you believe C;

Wr- If A, $B \models C$, then you have reason to see to it that if you believe A and believe B, you don't disbelieve C.

However, MacFarlane seems to be worried that the *Wr*'s are perhaps too weak, as they give only reasons - and not obligations - to believe the conclusion of an argument provided that one believes its premises. But, we might ask, is there a reason for not liking reasons? In a sense, reasons are more complicated to handle than permissions and obligations: unlike the traditional deontic operators, reasons come in degrees of strength, they often need to be weighed against each other, and their strength is dependent on the context. However, reasons can be applied to situations requiring norms that are more finely-grained than obligations and permissions, which explains why we could prefer reasons over other deontic operators in spite of their complexity.

6.3.3 Restall on logical norms

In addition to truth and falsity, some paraconsistent logics admit another truth-value, i.e. *both true and false*. According to these logics, that p is false doesn't preclude that

[12] I will leave aside the *-k* principles, i.e. the bridge principles where the agent knows or believes the antecedent, as I don't think they make much difference for the purposes of the discussion.

p be also true. The Liar Paradox sentence is a famous example of a sentence taking the value *both true and false* (propositions taking the third value are sometimes called 'gluts'). Therefore, from a paraconsistent point of view, believing that a proposition, p, is false doesn't imply that p is *disbelieved*. For this reason, whereas Broome and MacFarlane discuss logical norms concerning believing and disbelieving, Greg Restall explains informally how logical consequence is usually taken to constrain the *acceptance* and *rejection* of beliefs.

The relevant passage is the following:

> As we have seen, a one-premise, one-conclusion argument from A to B constrains acceptance/rejection by ruling out accepting A and rejecting B. This explanation of the grip of valid argument has the advantage of symmetry. A valid argument from A to B does not, except by force of habit, have to be read as *establishing* the conclusion. If the conclusion is unbelievable, then it could just as well be read as *undermining* the premise. (Restall (2005): 204)

A straightforward way to regiment the principle formulated informally in the passage is:

> If $A \vdash B$, you ought not both accept A and reject B.

The principle could be formulated as a wide-scoped norm (similarly to *Wo*-), i.e. $A \vdash B \Rightarrow O \neg (Acc(A) \wedge Rej(B))$, and I call this the *Consequence Principle* in the remainder of this chapter.

Now, it should be noticed that the logical norms presented differ substantially with respect to the deontic operator they include. We have seen that the normative requirements, *Wo*- and what I have called the Consequence Principle share an operator (*ought*) that is typical of *prescriptions* in a strict sense.

On the other hand, *Wr+* and *Wr*- share a deontic operator (*have reason to*) which brings about a weaker constraint. Following Broome (1999), we might call the latter constraints *recommendations*.

However, although *ought* is widely accepted as a deontic operator, it is more contentious whether *have reason to* can legitimately be deemed a deontic operator. The dubious status of reasons casts a general suspicion on recommendations. In addition to the general doubts on the status of recommendations, there are also specific worries about recommendations *as logical norms*. For instance, Broome (1999) argues that recommendations are too weak to capture the normative constraints imposed by logical consequence on thought.

In the last section of this chapter, the issue of the normativity of logic will be addressed directly. There, we will have enough material to make a general point about the normativity of logic, no matter what form they take in particular.

In the remainder, I consider two well-known epistemic paradoxes and put the logical norms presented to the test against them. By the end of the next section, it will become clear that the prescriptions in a strict sense and the recommendations deal in a very different way with the epistemic paradoxes.

6.4. ... and their problems

Before I set up the tests, there are some points that need to be set straight. First, the assumptions of our discussion in the remainder of the chapter need to be made explicit and, second, there are some distinctions that need to be highlighted.

There is an old debate between those who think that beliefs are an all-or-nothing sort of propositional attitude, and those who argue that beliefs come in degrees. According to the latter, our doxastic states vary in strength, and the thesis may seem intuitive as we experience having varying degrees of certainty about propositions.

However, there is a difference between the naïve view that we have degrees of belief, and the more radical view that beliefs respect the so-called *Lockean Thesis*, which has been discussed extensively by Foley (1993). The Lockean Thesis, which is our first assumption, can be summarized as:

> An agent, A, believes a proposition, p, if and only if A has subjective evidence that p is equal to or greater than a certain degree of belief (usually identified as 0.5 or greater).

Contrary to the naïve view on degrees of belief, the Lockean Thesis could be seen as having a *metaphysical* import, because it *defines* beliefs as states where an agent is confident about a proposition above a certain threshold level; in this sense, the Lockean Thesis reduces categorical (or qualitative) beliefs to credences (or quantitative beliefs).

However, as credences are often taken to conform to the constraints of probabilistic coherence, some (e.g. Frankish 2009) have objected that the Lockean Thesis needs to be rejected because it flies in the face of the Agglomeration Principle, or Closure under Conjunction. Agglomeration fails whenever believing a proposition, p, and believing another proposition, q, doesn't result in believing the conjunction $p \wedge q$. As it happens, having sufficient confidence in a proposition, p, and sufficient confidence in another proposition, q, doesn't guarantee that we have sufficient confidence in the conjunction, $p \wedge q$. Therefore, endorsing the Lockean Thesis rules out Agglomeration. To this criticism we can reply that, by the end of the section, it will become clear that there is no obvious reason why we should want to keep Agglomeration.

The second assumption is that I understand the statement 'There is at least one false sentence in the set of sentences, S' as being equivalent to 'The conjunction of all the sentences in S is false'.

Besides the assumptions I make in this chapter, I should also specify that the work presented in the current section is related to work done by Christensen (2004) on deductive cogency and epistemic rationality. Whereas Christensen's book addresses general constraints of deductive cogency, in the present chapter the epistemic paradoxes are used to test specific logical norms that are meant to be rationality requirements.

We would soon get into trouble if, referring to rationality requirements, we didn't specify what kind of rationality and rationality requirements we are talking about here. To this extent, Christensen (2004) comes in handy once again. Christensen[13] differentiates rationality into three orthogonal pairs: epistemic-pragmatic, synchronic-diachronic, and local-global. Christensen doesn't give definitions for each of the terms in the pairs, but he gives examples of how the corresponding forms of rationality could be manifested. Below, I try to sketch some informal definitions for each of the terms in the pairs, and I hope that I don't diverge too much from Christensen's insight.

Let's discuss the epistemic-pragmatic dichotomy first. Epistemic rationality is a standard prescribing that one's beliefs conform to some epistemic criteria and, most notably, the criterion would rather be the possession of evidence. On the other

[13] Christensen (2004): 4-11.

hand, pragmatic rationality aims at maximizing one's expected utility, as in Pascal's Wager.[14]

For example, suppose you are a detective conducting an investigation into a murder. Due to the scarcity of evidence, you cannot prove that the victim's wife committed the murder, although you are inclined to think that the wife would somehow benefit from her husband's death. Epistemic rationality doesn't allow that you believe that the wife murdered her husband.

But now, suppose that the police has been investigating the crime for a while without any tangible results. The good name of the local police department has been called into question by the press and by the public opinion. Then, the police chief starts to put pressure on you, asking you to find out who the killer is, with more and more insistency. According to pragmatic rationality, you might want to reveal the name of the victim's wife as the main suspect to soothe the general discontent with the investigations.

It is not uncommon to endorse that epistemic rationality can be reduced to instrumental rationality,[15] but for our purposes, it will suffice to make clear the distinction between the two.

Now, let's turn to the synchronic-diachronic dichotomy. Diachronic rationality governs the way beliefs are constrained over time (for example, it constrains belief revision), whereas synchronic rationality governs the way beliefs are constrained at a certain point in time (i.e. simultaneously).

Last comes the distinction between local and global rationality. The difference between the two is a matter of quantity. Whereas local rationality governs the way individual beliefs are connected to others, global rationality governs the way beliefs are connected to other beliefs in general or, in other words, holistically.

That said, let's spend a few words on what kinds of rationality we are going to deal with in our discussion: the rationality requirements imposed by logic on our beliefs are general and they don't have to do with our expected utility, so what we want to find out is whether they guarantee global and epistemic rationality.

On the other hand, the logical norms don't prescribe nor suggest how to revise our beliefs, therefore they are meant to preserve the synchronic rationality of our beliefs.

[14]Pragmatic rationality is often called *instrumental* rationality.

[15]For example, Nozick (1993) holds this view. See Kelly (2003) for a critical discussion of the reductionist view.

In the next subsection, we are going to look at the logical norms under the lens of epistemic, global and synchronic rationality.

6.4.1 Simone's manuscript

Consider the following Preface-paradox-like scenario.[16] Simone is about to publish a book on antibiotic-resistant bacteria, and she believes each of the propositions in her manuscript. However, due to her past experience, her awareness of being fallible, etc., Simone believes that there is at least one error in her manuscript.

The scenario tells us that Simone believes the propositions p_1, p_2, ..., p_i in her manuscript. By *Adjunction*, p_1, p_2, ..., $p_i \models p_1 \wedge p_2 \wedge ... \wedge p_i$.

However, Simone thinks that there is at least one error in her book, i.e. she also believes $\neg(p_1 \wedge p_2 \wedge ... \wedge p_i)$. It seems intuitively reasonable for Simone to believe that each of the propositions in her manuscript is true - she has run lab tests, she has strong evidence for each of the propositions, etc. It also seems intuitively reasonable for Simone to believe that there is at least one mistake in her book - she realizes that she has made mistakes in the past, she is aware that her colleagues' papers usually contain errors, etc.

Let's now examine how the logical norms deal with the Preface example. What I do in the remainder of the section is to consider normative requirements, MacFarlane's bridge principles, and the Consequence Principle in turn, to answer the question 'Is what the logical norms require Simone to believe *in fact* reasonable for her to believe?'. Answering this question will help determine whether the logical norms discussed meet their *desideratum.*

I start the Preface test considering how the logical norms fare in a full-belief scenario, then I will turn to discussing the same scenario in terms of credences. Moreover, as normative requirements and *Wo-* are both wide-scope obligations, I consider them together in my analysis.[17] According to the normative requirements and to *Wo-*, Simone is expected to meet the following requirement:

---

[16]The Preface Paradox was discussed for the first time by Makinson (1965).

[17]Normative requirements are quite charitably treated the same way as *Wo-*. In fact, Broome (1999) describes the requirements in a way that actually justifies their translation as instances of *Wo+* (*Wo+* reads 'If A, $B \models C$, then you ought to see to it that if you believe A and believe B, you believe C'), with the proviso that the antecedent, i.e. 'If A, $B \models C$', is obvious. In general, *Wo+* fails to be a plausible principle because ordinary agents lack logical omniscience. In the case of Broome's normative requirements, I don't think that the entailment relation being obvious makes much difference with respect to the Preface and the Lottery paradoxes.

> *Normative requirements, Wo-*: Simone ought to see to it that if she believes p_1, p_2, ..., p_i, she doesn't disbelieve $p_1 \wedge p_2 \wedge ... \wedge p_i$.

However, Simone believes that there is at least one error in her manuscript, i.e. Simone disbelieves $p_1 \wedge p_2 \wedge ... \wedge p_i$, so *Wo-* and the normative requirements impose an obligation that would be unreasonable for Simone to meet. So, *Wo-* and normative requirements don't meet the *desideratum* of logical norms in general, because they don't connect a valid argument, p_1, p_2, ..., $p_i \models p_1 \wedge p_2 \wedge ... \wedge p_i$, and informal reasoning in a plausible way.

Now, I examine the Consequence Principle and determine whether it fares any better, providing a reasonable requirement to Simone.

According to my formulation of the Consequence Principle, we should expect the following requirement to be enforced:

> *Consequence Principle*: Simone ought not both accept p_1, p_2, ..., p_i, and reject $p_1 \wedge p_2 \wedge ... \wedge p_i$.

Upon replacing the pair *accept-reject* with the pair *believe-disbelieve*, the Consequence Principle looks much like *Wo-*. The requirement that the principle imposes seems to fare like *Wo-* and Broome's normative requirements, too - it prohibits Simone to accept each of the propositions in her book and accept that there is at least an error somewhere in her manuscript (i.e. reject $p_1 \wedge p_2 \wedge ... \wedge p_i$). However, it seems reasonable that Simone believes each of the propositions in her manuscript and still refuses to believe their conjunction. So, similarly to *Wo-*, the Consequence Principle doesn't satisfy the general *desideratum* of logical norms.

Now, we have only two principles left to examine: *Wr+* and *Wr-*. According to the *Wr*'s, Simone's case would be subject to the following norms:

> *Wr+*: Simone has reason to see to it that if she believes p_1, p_2, ..., p_i, she believes $p_1 \wedge p_2 \wedge ... \wedge p_i$;
> *Wr-*: Simone has reason to see to it that if she believes p_1, p_2, ..., p_i, she doesn't disbelieve $p_1 \wedge p_2 \wedge ... \wedge p_i$.

The *Wr*'s don't impose an obligation to believe each of the propositions and their conjunction - they just say that Simone has a *pro tanto* reason to see that if she believes each of the propositions in her book, she believes (doesn't disbelieve) their conjunction. Therefore, the *Wr*'s are not threatened by there being other epistemic norms, such as epistemic modesty, recommending an opposite course of action, so they fare better than *Wo*- and the normative requirements.

However, it might be objected that the *Wr*'s don't have the belief-constraining power that we would expect from a logical norm. For example, Broome (1999) argues that, provided that $p \models q$, 'if you do believe p and yet do not believe q, you are not entirely as you ought to be. So the relation is strict, which means it cannot be the reasons relation.' (Broome (1999): 403). For the purpose of this chapter, I don't delve into the debate on the *pros* and *cons* of reasons. For now, I just acknowledge that there are concerns about reasons being not sufficiently strong to describe the normativity of logic on thought.

Let's consider now the possible responses to the Preface example. A quite immediate reply to the Preface paradox is that it is not a real paradox. In fact, it could be solved separating first-order beliefs from second-order beliefs. In our example, the first-order beliefs are Simone's beliefs in each of the propositions in her manuscript, whereas the second-order belief is Simone's modest claim that there is at least one error in her book.

However, as Christensen (2004) puts it, the point about first- and second-order beliefs shouldn't be taken as conclusive against the Preface case:

> This line seems unpromising to me. For it seems clear that an author who knew what she had said in the body of her book could realize that this conjunction was materially equivalent to the second-order claim of inerrancy for the body of the book. Once she has accepted the equivalence, closure will take her from the conjunction to the second-order claim. (Christensen (2004): 37)

Just consider the conjunction of all the propositions in Simone's book, $p_1 \wedge p_2 \wedge ... \wedge p_i$. If Simone believes the conjunction of the propositions in her book, she *must* believe that her book contains only true statements, which is already a second-order belief. In fact, it would seem unnatural that Simone could believe $p_1 \wedge p_2 \wedge ... \wedge p_i$ and not believe that her book is error-free.

The inconsistency between Simone's beliefs still remains. What other candidates could be dismissed, then?

So far, we have found that the logical norms in trouble with the Preface Paradox are *Wo*-, the normative requirements, and the Consequence Principle. Therefore, one option could be to dismiss the troublesome logical norms.

There is another option, though. One might rather decide to dismiss the *Agglomeration Principle* (also called *Closure under Conjunction*), i.e. $B\varphi$, $B\psi \rightarrow B(\varphi \wedge \psi)$. The Agglomeration Principle can be read as a descriptive principle or as a normative principle. As a descriptive principle, Agglomeration says that whenever an agent believes proposition φ and believes ψ, she believes $\varphi \wedge \psi$. The descriptive reading of the Agglomeration Principle, besides not being very interesting from an epistemological point of view, doesn't seem very plausible. On the other hand, the normative reading of Agglomeration says that rationality requires that if an agent believes φ and believes ψ, she believes $\varphi \wedge \psi$. Following Broome's general analysis of requirements, I formalize Agglomeration as a wide-scoped obligation: $O((B\varphi \wedge B\psi) \rightarrow B(\varphi \wedge \psi))$.

Now, it is relatively uncontroversial that the descriptive version of Agglomeration fails, but it might not be as uncontroversial that the normative version of Agglomeration fails as well. Therefore, if our aim is to rescue the logical norms in trouble, a possible reply to the Preface Paradox could be to give up the normative reading of Agglomeration.

Furthermore, this option seems to be supported by probabilism. Suppose that Simone's manuscript is conveniently constituted of a hundred propositions, p_1, ..., p_{100}, and that Simone has strong evidence for each of them. To keep the math simple, let's say that she believes each of the propositions to degree 0.99 and that the propositions are independent. Consequently, according to Bayesianism, given that the probability of the conjunction of independent propositions is given by the product of their individual probabilities, the credence with respect to $p_1 \wedge ... \wedge p_{100}$ is $0.99^{100} = 0.366$[18] - well below the threshold for accepting $p_1 \wedge ... \wedge p_{100}$.

However, let's concede that it is not straightforward that having a credence of 0.366 with respect to $p_1 \wedge ... \wedge p_{100}$ implies that one disbelieves $p_1 \wedge ... \wedge p_{100}$. For example, consider a Lockean belief system with a rather high threshold for belief, i.e. we

[18]For the sake of simplicity, I am assuming that the propositions are independent.

say that Simone believes a proposition, p, whenever her credence that p is equal or greater than 0.7. As Simone's credence with respect to $p_1 \wedge \ldots \wedge p_{100}$ is 0.366, her credence with respect to $\neg(p_1 \wedge \ldots \wedge p_{100})$ is 0.634, which is below the threshold. In such a case, giving up Agglomeration doesn't commit one to reject *Wo*- nor any of the other logical norms in trouble.[19]

I will show that giving up the Agglomeration Principle is not enough to save the logical norms in trouble, though. In the next section, I use a probabilistic solution to the Lottery Paradox to show why this is the case.

6.4.2 The lottery

In the following, I set up a typical Lottery Paradox scenario, then I consider how each of the logical norms deal with it.[20]

Suppose that a fair lottery has 1,000 tickets and exactly one winning ticket among them. It is rational to believe that some ticket will be the winning one.

Now, suppose that you have a ticket from the lottery. Given the odds, it seems perfectly reasonable for you to believe that it is very unlikely that you have the winning ticket.

Let p_1 stand for 'This is the winning ticket' and $\neg p_1$ stand for 'It is not the case that this is the winning ticket'.

Then, assuming a Lockean view, as the odds that your ticket is the winning one are very low (1/1,000), you are justified to believe $\neg p_1$.

In fact, given that there is exactly one winning ticket out of a thousand, it is reasonable for you to also believe $\neg p_2$, ..., $\neg p_{1000}$. By *Adjunction*, we have that $\neg p_1, \neg p_2$, ..., $\neg p_{1000} \models \neg p_1 \wedge \neg p_2 \wedge \ldots \wedge \neg p_{1000}$ and, by De Morgan's Law, $\neg p_1 \wedge \neg p_2 \wedge \ldots \wedge \neg p_{1000}$ is equivalent to $\neg(p_1 \vee p_2 \vee \ldots \vee p_{1000})$, so $\neg p_1, \neg p_2$, ..., $\neg p_{1000}$ entails that there is no winning ticket.

Now, let's examine again each of the logical norms to determine whether what they require you to believe is in fact reasonable for you to believe.

[19] If we set up a Preface scenario where the book contains 1,000 propositions, Simone would have a much lower credence in their conjunction (indeed, it would approximate to 0) and, in turn, Simone would have a credence in the negation of the conjunction that largely exceeds the threshold. Therefore, we could well claim that Simone disbelieves the conjunction of all the propositions in her manuscript. However, I want to show that even conceding a favourable case to our opponent won't help their view.

[20] The Lottery Paradox was first discussed in Kyburg (1961).

According to the normative requirements and *Wo-*, you are expected to meet the following requirement:

> *Normative requirements*, *Wo-*: You ought to see to it that if you believe $\neg p_1, \neg p_2$, ..., $\neg p_{1000}$, you don't disbelieve $\neg(p_1 \vee p_2 \vee ... \vee p_{1000})$.

However, you do know that *some* ticket is the winning one. So, as in the Preface Paradox case, *Wo-* and the normative requirements don't seem to prescribe a reasonable epistemic requirement, so failing to meet the main *desideratum* of logical norms.

The situation is not very different with the Consequence Principle, which gives the following requirement:

> *Consequence Principle*: You ought not both accept $\neg p_1, \neg p_2$, ..., $\neg p_{1000}$, and reject $\neg(p_1 \vee p_2 \vee ... \vee p_{1000})$.

The principle seems to fare like *Wo-* and Broome's normative requirements in the Lottery case, too. In fact, the Consequence Principle prohibits you to accept $\neg p_1, \neg p_2$, ..., $\neg p_{1000}$ and reject that there is no winning ticket, which doesn't seem reasonable.

On the other hand, *Wr+* and *Wr-* impose slacker constraints in the Lottery case:

> *Wr+*: You have reason to see to it that if you believe $\neg p_1, \neg p_2$, ..., $\neg p_{1000}$, you believe $\neg(p_1 \vee p_2 \vee ... \vee p_{1000})$;
> *Wr-*: You have reason to see to it that if you believe $\neg p_1, \neg p_2$, ..., $\neg p_{1000}$, you don't disbelieve $\neg(p_1 \vee p_2 \vee ... \vee p_{1000})$.

The *Wr*'s don't impose an obligation to believe each of the negated propositions and to believe that there is no winning ticket - they just say that you have a *pro tanto* reason to see that if you believe each of $\neg p_1, \neg p_2$, ..., $\neg p_{1000}$, you believe (don't disbelieve) that there is no winning ticket. As such, the *Wr*'s fare better than *Wo-* and the normative requirements. As with the Preface case, it might be objected that reasons don't have as much deontic force as we would expect from a logical norm. Again, I leave this issue open for the moment.

From the Lottery Paradox test, it turns out that *Wo*-, the normative requirements and the Consequence Principle fail to meet the general *desideratum* of logical norms, because they don't express plausible ways to connect valid arguments and informal reasoning. On the other hand, *Wr+* and *Wr*- satisfy the *desideratum* but they impose much weaker constraints on belief.

Again, two paths can be pursued to deal with the Lottery test. Similarly to the Preface case, the first option is to give up the logical norms in trouble.

The second option - i.e. rejecting Agglomeration - could *prima facie* seem a feasible way out in the Lottery case, too. One might be tempted to support the choice of dismissing Agglomeration by appealing to probabilism again. However, this time resorting to probabilism commits one to rejecting more than just the Agglomeration Principle.

Consider a weaker version of the Agglomeration Principle, which I call *Weak Agglomeration*:

$$\textit{Weak Agglomeration: } O(B\varphi,\, B\psi \rightarrow \neg \textit{Dis}(\varphi \wedge \psi)).$$

The Lottery case appears to be a counterexample even to Weak Agglomeration. In fact, as we have seen, given the odds of the lottery it is reasonable for you to believe an instance of the proposition 'This is not the winning ticket' for each of the tickets from the raffle. However, you know that the lottery is fair and that there is exactly one winning ticket, therefore it is equally reasonable for you to reject that there are no winning tickets.

Similarly to the Preface Paradox, Bayesianism doesn't help to rescue the logical norms in the Lottery case. To see the point, consider the following solution to the Lottery Paradox:

$p_1, p_2, ..., p_{1000}$ are not mutually independent, as ticket 1, ticket 2, ..., ticket 1000 are tickets from the same lottery and the probability of $(p_1 \vee p_2 \vee ... \vee p_{1000}) \vee \neg(p_1 \vee p_2 \vee ... \vee p_{1000})$ is equal to 1. In fact, for the principles of probabilism, any logical tautology gets assigned value 1 as, according to probabilism, the probability of a disjunction, $\varphi \vee \psi$, with mutually inconsistent disjuncts is equal to the sum of the disjuncts' probabilities. In other words, $p(\varphi \vee \psi) = p(\varphi) + p(\psi)$.

So, we know that the probability of $(p_1 \vee p_2 \vee ... \vee p_{1000}) \vee \neg(p_1 \vee p_2 \vee ... \vee p_{1000})$ is equal to 1; therefore, $p(p_1 \vee p_2 \vee ... \vee p_{1000}) + p(\neg(p_1 \vee p_2 \vee ... \vee p_{1000})) = 1$.[21]

But we also know that there is exactly one winning ticket, i.e. we know that the probability of $(p_1 \vee p_2 \vee ... \vee p_{1000})$ is equal to 1. Therefore, the probability of $\neg(p_1 \vee p_2 \vee ... \vee p_{1000})$ (and indeed, of $\neg p_1 \wedge \neg p_2 \wedge ... \wedge \neg p_{1000}$) must be 1 - 1 = 0. Put in probabilistic jargon, $\neg p_1, \neg p_2, ..., \neg p_{1000}$ *exhaust* the probabilistic space.

Now, let's go back to the Preface example. A quick look at Bayesianism had shown that the probability of the conjunction of all the propositions in Simone's manuscript (which was 0.366, as we had seen) cannot guarantee that Simone believes $p_1 \wedge p_2 \wedge ... \wedge p_{100}$. However, we have seen how the propositional attitude of not-believing (or failing to believe) a proposition, p, is not the same as the propositional attitude of disbelieving a proposition, p. We conceded that Simone not believing $p_1 \wedge p_2 \wedge ... \wedge p_{100}$ doesn't mean that Simone disbelieves $p_1 \wedge p_2 \wedge ... \wedge p_{100}$, provided our Lockean belief system has a threshold of at least 0.7. As far as the Preface case is concerned, Bayesianism really seems to reconcile the need to reject Agglomeration with the need to keep the prescriptive logical norms.

However, the Bayesian solution to the Lottery Paradox shows that rejecting the Agglomeration Principle is not sufficient to save the principles in trouble. This time, the solution to the Lottery says, the probability for $\neg(p_1 \vee p_2 \vee ... \vee p_{1000})$ - i.e. the probability that there is no winning ticket - is 0. In other words, it must be plainly false that no ticket will win the lottery.[22] Therefore, this time you are perfectly justified in disbelieving (or rejecting) that no ticket will win the lottery. But this is exactly the opposite of what the logical norms (except *Wr+*, *Wr-*) prescribe to you to do.

In fact, as we have seen, *Wo-*, normative requirements, and the Consequence Principle require you to see to it that if you believe $\neg p_1, \neg p_2, ..., \neg p_{1000}$, you don't disbelieve/reject $\neg p_1 \wedge \neg p_2 \wedge ... \wedge \neg p_{1000}$. But here the logical norms seem to conflict with probabilism, because probabilism requires that you assign probability 0 to $\neg p_1 \wedge \neg p_2 \wedge ... \wedge \neg p_{1000}$.

Notice that the normative reading of Weak Agglomeration is just an instance of *Wo-*. Therefore, the option of rejecting Weak Agglomeration and retaining *Wo-* is closed

[21]The notation used here might be confusing. The occurrences of p before the parentheses denote 'the degree of probability of', whereas the occurrences of p inside the parentheses denote propositions.

[22]Here, I am assuming that the agent *knows* that there is exactly one winning ticket.

off now, showing that there is no immediate way we can get away with rejecting Agglomeration and Weak Agglomeration to rescue the logical norms in trouble.

Some, e.g. Kolodny (2005), oppose the view that rationality requirements have wide scope. For the sake of the argument, let's consider the possibility that Agglomeration is actually a narrow-scope requirement, i.e. $(B\varphi \wedge B\psi) \rightarrow OB(\varphi \wedge \psi))$.

Narrow-scope Agglomeration says that, for instance, if you believe $\neg p_1, \neg p_2$, ..., $\neg p_{1000}$, you ought to believe that there is no winning ticket. We can narrow down the scope also in Weak Agglomeration, so that it becomes $(B\varphi \wedge B\psi) \rightarrow O\neg Dis(\varphi \wedge \psi))$. Adapted to the Lottery example, narrow-scope Weak Agglomeration says that if you believe $\neg p_1, \neg p_2$, ..., $\neg p_{1000}$, you ought not disbelieve that there is no winning ticket.

Still, as $\neg(p_1 \vee p_2 \vee ... \vee p_{1000})$ takes value 0, you have sufficient grounds to rationally disbelieve that no ticket will win the lottery. Therefore, narrowing down the scope of Agglomeration and Weak Agglomeration, respectively, doesn't make them immune to the counterexample.[23]

There is another possible way out of the stalemate that we haven't explored, yet. It is a much more radical solution than rejecting Agglomeration, nonetheless it is worth considering it. Instead of rejecting Agglomeration, which is the epistemic counterpart of Adjunction, we could decide to reject Adjunction. Dismissing Adjunction, we would prevent the Preface Paradox.

In fact, consider once again Simone's manuscript, consisting of a finite number of propositions, p_1, p_2, ..., p_i. If Adjunction doesn't hold, we would have that p_1, p_2, ..., $p_i \nvDash p_1 \wedge p_2 \wedge ... \wedge p_i$. Therefore, Simone's belief that there is at least one mistake in her manuscript, which we translated as $\neg(p_1 \wedge p_2 \wedge ... \wedge p_i)$, wouldn't be inconsistent with the conjunction $p_1 \wedge p_2 \wedge ... \wedge p_i$, classically entailed by Simone's belief in each of the propositions in her book.

Clearly, rejecting Adjunction would also stop the Lottery Paradox. There, we saw that considering that each of the ticket, in turn, is hardly going to win the lottery, $\neg p_1, \neg p_2$, ..., $\neg p_{1000}$, entails that no ticket will win the lottery, $\neg p_1 \wedge \neg p_2 \wedge \neg p_{1000}$. However, rejecting Adjunction, we have that $\neg p_1, \neg p_2$, ..., $\neg p_{1000} \nvDash \neg p_1 \wedge \neg p_2 \wedge \neg p_{1000}$. Therefore, our belief that there is some winning ticket, $p_1 \vee p_2 \vee ... \vee p_{1000}$, doesn't bring about any paradox.

Hence, it looks like rejecting Adjunction might be the way to go to cope with the epistemic paradoxes. The examples of non-adjunctive logic that come to mind first

[23]There are also independent concerns with narrow-scoped principles, as I mentioned.

are Rescher and Manor's paraconsistent non-adjunctive logic,[24] which allows to conjoin the premises unless they are inconsistent. But, although non-adjunctive logics are effective in blocking the paradoxes, they are subject to the independent worries common to paraconsistent logics, which would require thorough discussion, so we won't consider the option of rejecting Adjunction in this chapter.

In this chapter, we have seen that logical norms, similarly to epistemic norms, are meant to maximize rationality. However, logical and epistemic norms are sometimes in tension, and following logical norms alone doesn't guarantee that the resulting beliefs will be rational. For the purpose of this chapter, our aim is to prove that, given certain assumptions, some logical norms are not plausible, as they prescribe that we believe what we shouldn't rationally believe, e.g. that the Lottery doesn't have a winning ticket. In particular, in this section I have shown that *normative requirements*, *Wo-*, and the *Consequence Principle* don't withstand the Preface nor the Lottery case. *Wr+* and *Wr-* fare better than the other norms but there is concern that they might be too weak. Therefore, if we concede that there are credences (and something like the Lockean Thesis), we cannot also have prescriptive logical norms.

In the next section, I try to determine what category of normativity logic falls under. I draw from the taxonomy of norms discussed in Glüer and Wikforss (2010) and then I determine under what category the normative role of logic can be classified.

6.5. Types of normativity

As we have seen in the previous sections, logical norms are a very peculiar kind of norms. They can come in different varieties and normative strength, according to the deontic operators they employ, the scope of the deontic operator, etc.

After our long discussion of the logical norms, the reader might have lost track of all the subdivisions norms are classified into. To help the reader remember which labels means what, here is a summary of the classification of norms according to Glüer and Wikforss (2010).

1.: *Norms for action*: these norms are about what to do. Legal constraints are an example of such norms;

[24]See Rescher and Manor (1971)

2.: *Norms of being*: these norms are similar to evaluations, i.e. they are about what state of affairs are valuable and ought to obtain.

Among the norms for action, there are several subcategories:

1.1: *Instrumental norms*: these norms are contingent upon the agent's goal. For example, 'I should run every day to be able to run the London marathon';
1.2: *Non-instrumental norms*: these norms are not contingent upon the agent's goal. Moral constraints are usually taken to exemplify this category of norms, such as 'You should always help a person in need';
1.3: *Prima facie norms*: these rules can be overridden by other rules. For example, 'You shouldn't overtake a car on its left' (assuming that you drive in the UK);
1.4: *Categorical norms*: these rules can't be overridden.

Non-instrumental norms can be divided into subcategories as well:

1.2.1: *Prescriptions*: these norms can be formulated with an *ought*;
1.2.2: *Constitutive rules*: in these norms, the action subject to the norm owes its existence to the norm. The rules of games are an example of such norms.

Prescriptions can be:

1.2.1.1: *Conditional*, such as 'If someone holds the door for you, you should thank them';
1.2.1.2: *Unconditional*, such as 'You should donate to charity'.

In addition to this, the deontic operator in the conditional prescriptions can have wide scope or narrow scope.

As we have seen in the previous sections, logical norms are a very peculiar kind of norms: they can come in different varieties and normative strength, according to the deontic operators they bear, the scope of the deontic operator, etc. Logical norms are often formulated as norms for action. In fact, they are supposed to regulate how we ought, have reason to or, at least, how we shouldn't, reason.

In turn, according to the taxonomy, norms for action can be instrumental or non-instrumental. As logical norms are not contingent upon the agent's goals, they seem to fall under the latter category. However, there is an important sense in which the distinction between instrumental and non-instrumental norms as it is formulated in Glüer and Wikforss (2010) doesn't do justice to logical norms.

To see this, I consider the distinction between the *evaluative* and the *directive* sense of *ought* introduced by Thomson (2008). As Elqayam and Evans (2011) have pointed out, there is a third sense of *ought*, i.e. the *functional* sense. The functional *ought* is relative to - among other things - biological functions, as in 'Heart ought to pump the blood', and doesn't have a deontic character. Functional *ought*'s just describe what the proper function of an entity (such as an organ) should be.

Unlike the functional use of *ought*, the evaluative and the directive *ought's* have a deontic meaning. The evaluative *ought* is appropriate to express moral claims, such as 'You ought to tell the truth', whereas the directive *ought* is meant to recommend a certain course of action to reach an end, as in 'You ought to fold a sheet this way' (supposing you want to make a paper plane, for example). The latter example especially shows that directive *ought*'s are easily thought of as having an instrumental meaning. As Elqayam and Evans (2011) acknowledge, the directive and evaluative meanings of *ought* are often mixed in common discourse. Indeed, many moral claims could be seen as including a mix of directive and evaluative *ought*'s.

Although the difference between the directive *ought* and Glüer & Wikforss' [25] *instrumental norms* doesn't shine through the example 'You ought to fold a sheet this way', it is important to see that a directive *ought* is not necessarily contingent upon the immediate agent's goal.

Consider the example 'You ought to donate to Oxfam'. In this example, the *ought* is evaluative because it is used to express that the action of donating to Oxfam is valuable. However, the action of donating to Oxfam is not simply valuable in itself. It is valuable because donating to Oxfam is supposed to be a means to, e.g. helping people living in poverty around the world. Therefore, the use of *ought* in 'You ought

[25] Which I will call *G&W* in the following.

to donate to Oxfam' doesn't seem merely evaluative, but it seems to be directive as well.

The evaluative+directive use of *ought* is not accounted for by G&W, but I think that their taxonomy could be fruitfully supplemented with the categories of the directive and the evaluative *ought*, which are not meant to be mutually exclusive.

Now, to go back to our question, logical norms are not instrumental according to the taxonomy outlined by Glüer and Wikforss (2010), but the logical norms I have discussed in the previous sections are recognizable as directive norms nonetheless. In fact, logical norms are meant to be epistemic norms aimed at maximizing the rationality of our beliefs. For this reason, I think that portraying logical norms as *purely* evaluative doesn't do justice to their purpose. It is very plausible that logical norms could be best located under the category evaluative+directive.

Going back to G&W's taxonomy, the third category divides norms into prescriptions and constitutive norms.

Now, if logical norms were constitutive of thought, we would have no mental process identifiable as proper thought outside them. Although a similar view could perhaps be attributed to Kant[26] and Frege,[27] the idea that agents need to comply with logical laws if they are going to think at all seems far from reality. I hope to have shown in the previous section that logical norms are not always successful in imposing an obligation about what is rational to believe.

There is a last thing to point out. We have seen that we can draw an initial distinction between norms that are prescriptive in a strict sense, and recommendations, and we have also seen that G&W appear to use the label 'prescriptive' in a more general way. In fact, G&W understand prescriptive norms to be norms that employ deontic vocabulary. In other words, prescriptive norms can be formulated as obligations, prohibitions, or permissions. Although reasons are not taken into account in G&W's

[26]Kant thought that we cannot dispense from thinking logically, if we are to think at all. However, it should be kept in mind that Kant's idea of logic doesn't have much to do with the current idea of logic as a system consisting in a logical language, language-forming rules and inference rules (or axioms, if so preferred).

[27]In his introduction to *Grundgesetze*, Frege writes: 'Every law stating what is the case can be conceived as prescriptive, one should think in accordance with it, and in that sense it is accordingly a law of thought. This holds for geometrical and physical laws no less than for the logical. The latter better deserve the title "laws of thought" only if thereby it is supposed to be said that they are the most general laws, prescribing how to think wherever there is thinking at all.' (Frege (2013): xv)

definition of prescriptive norms, it is really up to us to decide whether reasons deserve a place in the deontic vocabulary or not.

Certainly, logical norms can be formulated as requirements, i.e. obligations. However, the examples with the epistemic paradoxes have shown that such logical norms violate epistemic, global and synchronic rationality in some circumstances. Therefore, if we want logical norms to be rationality-preserving principles, we'd better not formulate them as obligations.

If reasons are not admitted in the deontic vocabulary, it is not clear what normative category logical norms fall under. They are not constitutive norms, but they are not prescriptive either. If reasons enter the deontic vocabulary, we could say that logical norms are prescriptive in a broad sense, as they specify recommendations (reasons) for belief.

If my analysis is correct, this implies three claims: (i) logical norms can be overridden by other norms, giving us reason to put logical norms under the category of *prima facie* norms; (ii) classifying logical norms as *prima facie* rules out that they can be constitutive norms; (iii) in virtue of being non-constitutive, logical norms fall under the category of prescriptive norms.

6.6. Conclusion

The logical norms that I have discussed in this chapter can look *prima facie* as an attractive way to formulate the bulk of a strong theory of the normativity of logic. Such a theory has it that conforming our beliefs to logical consequence constitutes a necessary requirement for epistemic rationality.

However, we have seen that the epistemic paradoxes threaten a strong approach to the normativity of logic. In particular, we have seen that the logical norms cannot accommodate our intuitions about what a rational agent *ought* to believe in cases such as the Preface Paradox and the Lottery Paradox. The logical norms formulated as obligations (*Wo*-, normative requirements, and the Consequence Principle) can only be held as *prima facie* norms. In other words, the attempts to provide logical norms as necessary requirements for epistemic rationality are not successful.

We have seen that sometimes we judge the most rational course of action to be that recommended by an epistemic norm, rather than that recommended by a logical norm. For this reason, the logical norms formulated as recommendations (*Wr*+, *Wr*-) capture better the normativity of logic, because they leave room for other rationality requirements and epistemic norms to be considered.

On the other hand, reasons are subject to some worries themselves. For instance, some may object that reasons fail to provide logical norms that have the belief-constraining power that we would expect from logical norms. Others may be concerned about reasons not being real deontic operators. Therefore my conclusion, for those who endorse these worries about reasons, is that none of the logical norms presented is plausible at all.

Conclusion

Let's consider the questions that opened the thesis once again: What normative import does logic have on our reasoning? How do, and how should, logical norms influence our rational thought? We can now answer these questions as follows: logic has a weak influence on our reasoning, which resembles the normative power of recommendations rather than that of obligations. To the second question, we can reply that logical norms do not and should not always influence our thought in order for it to be rational.

We have seen how this change in the perspective on the normativity of logic has been the effect of other changes, namely the development of non-classical logics and, especially, the realistic turn in the theories of rationality, which could be assimilated to Copernican revolutions in their respective fields.

Philosophers are interested, among other things, in better understanding ourselves, and the study of rationality from the point of view of real agents is a valuable way of doing that. In theory, logic is what we have studied in our philosophy courses, or what Peirce called *logica docens*, and it is (we like to think) infallible and invincible; in practice, sometimes logic conflicts with independent epistemic requirements and loses the match. I hope to have shown that this is something we shouldn't be too worried about.

The main characters of my thesis are logic, normativity, and rationality, whose interplay defines the normativity of logic. For this reason, I devoted a chapter of my thesis to each of them in order to prepare the ground for the discussion of logical normativity in the final chapter. Let me sum up the chapters and how they contributed to the thesis.

The first chapter lays the basis for the rest of the thesis, focusing on the definitions of logical validity and their problems. In addition, the chapter explores a non-traditional account of validity, and rejects it. The second chapter asks whether logic is a priori and whether it provides a priori norms for reasoning, arguing that logic does provide norms for reasoning we are a priori entitled to apply. The third chapter delves into one of the most detailed accounts of the normativity of logic currently on the market, laying the basis for the final chapter. The fourth chapter addresses some theories of rationality that focus on real agents, rather than on ideal agents; I assume the characterization of 'rationality' provided by the theories discussed in the fourth chapter in my analysis of logical normativity in the final chapter. The fifth chapter discusses the main results in the psychology of reasoning, which demonstrate that the theory of rationality discussed in the fourth chapter is very plausible. Finally, the sixth and final chapter goes back to the issue of the normativity of logic. There, I ask whether there are any logical norms that provide a plausible way to connect formal logic to informal reasoning, i.e. reasoning that is rational in the sense discussed in the fourth chapter.

In detail, in the first chapter, I argued against primitivism about validity. First, we saw that validity can be defined in different ways and it can even be left undefined, as in Field (forthcoming). The competing accounts of validity are the model-theoretic, the proof-theoretic, the deflationist, and what we called the primitivist account of validity, which has been defended by Field. Every theory of validity suffers from particular problems but, in the first chapter, I addressed specifically the primitivist account because it has the claim that because of the problems besetting all other accounts, it has to be right in not burdening validity with an analysis that is doomed to fail.

I argued that, given all the considerations, the primitivist account is not in a better position than the other accounts of validity. The primitivist's formula was found to return results outside the [0, 1] interval, and I argued that the account raises serious concerns, one of which is common to the model-theoretic view of validity. Then, I suggested a definitional approach to validity very similar in spirit to Field's and found that it didn't appear to fare worse than the primitivist account of validity. I concluded that, if there are no strong reasons to endorse a justification reductionist account of validity, neither are there to prefer (FV) over the model- and proof-theoretic accounts of validity.

In the second chapter, I asked whether logic is a priori, and concluded that a fallibilist apriorism is the most plausible option. The traditional definition of *a priori knowledge* claims that some proposition is known a priori iff it is known (or justified) independently of experience. However, we saw that the definition raised questions about what 'indepedently of experience' means. I first addressed the role of a posteriori evidence on a priori knowledge, and I surveyed some positions on the relevance of empirical evidence to the a priori, explaining what characteristics empirical evidence should have to deprive the knowledge of its a priori status. Next, I introduced Putnam's empiricism and I summarized the main replies to it. Subsequently, I discussed the positive and the negative accounts of the a priori. The positive definitions of the a priori specify what is needed in the entitlement of a proposition, p, in order for p to be known a priori. In contrast, the negative characterizations of the a priori specify what needs to be absent from the entitlement to believe p, for p to be known a priori. I discussed the negative accounts first, focusing especially on Field's apriorism, his fallibilist view on the a priori, and the thesis that we have a default entitlement to apply logical norms. Next, I addressed the positive characterizations of the a priori: (i) as knowledge gained through rational intuition, as discussed e.g. by BonJour and by Bealer, and (ii) as knowledge that is true and justified in virtue of language alone, as discussed e.g. by Boghossian. Subsequently, we focused on the debate between the empiricist fallibilism promoted by Bueno and the aprioristic fallibilism endorsed by Field, and I argued that Bueno's criticism of Field's apriorism are not conclusive. In the last part of the chapter, I examined Williamson's view on the apriori, which holds that the distinction between a priori and a posteriori is not philosophically meaningful. In response to Williamson's argument, I rejected the main assumption on which it relies. I concluded that the only position which bears scrutiny is Field's aprioristic fallibilism, which holds that logical rules enjoy a default justification and that we are entitled by default to use them.

In the third chapter, I addressed the most promising of all current theories of the normativity of logic, i.e. MacFarlane's, which is particularly relevant to the last chapter of the thesis. I first discussed the interplay between the formality of logic, which MacFarlane addresses in his doctoral dissertation, and the issue of the normativity of logic, which MacFarlane addresses in a more recent paper. Then, I discussed the bridge principles introduced in MacFarlane (2004), agreeing with MacFarlane that some of them (dubbed *Wr+, Wr-,* and *Wo-*) describe *prima facie* plausible ways of framing the constraints imposed by logic on informal reasoning. Next, I analyzed the deontic operators used in the bridge principles, distinguished five senses of *ought* and made a hypothesis about the sense in which *ought* is employed in *Wo-*. Finally, I argued that one of the operators included in the bridge principles (i.e. *has reason to*) could be made sense of through the minimal deontic logic developed in Chellas (1974, 1980).

In the fourth chapter, I discussed at length the theories of rationality that provide the theoretical background for the later chapters (Cherniak (1986) and Harman (1986)), which I called the *minimal rationality* views. They focus on the study of rationality with regard to real agents; both the approaches take into account the cognitive, temporal and memory limitations of human agents. We saw that, although intuitively appealing, the distinction between proof (or argument) and reasoning was not succesfully shown by Harman's argument. In spite of this, I kept the distinction as a significant one, since it addresses two different mental processes that are crucial for the debate on the normativity of logic. The last section of the chapter introduced Peirce's account of *logica docens* and *logica utens*, and I suggested that *logica utens* could be equivalent to what the psychology of reasoning has called *mental logic*.

The fifth chapter focused on the psychology of reasoning. We saw that until the late 1960's, the mainstream view of the psychology of reasoning held that human reasoning follows roughly the same principles as formal logic (often referred to as Mental Logic Theory). In 1966, Peter Wason developed his famous selection task to test people's deductive reasoning skills, and the results of his test became a knock-down counterexample to Mental Logic. After discussing the selection task, I distinguished the main reactions to the experimental results and focused on a particular strategy to rescue the core idea underlying Mental Logic, which I called the *Wrong Conditional Reply*. We saw that the alternative versions of the selection task with manipulated semantic content provide evidence against the Wrong Conditional Reply. Then, I discussed the main theories of reasoning and how they explain the experimental results. In the last section, I argued that if we agree that the selection task and its alternative versions are valuable psychological evidence, we should conclude either that deductive reasoning is not indispensable for rationality, or that the responses given by people to the task show that they are often *arational*, rather than irrational.

In the sixth and final chapter, I addressed the normativity of logic, discussing some principles presented by Broome (1999), MacFarlane (2004), and Restall (2005). We saw that the main *desideratum* of the principles is that they provide a plausible way to explain the connection between valid arguments and informal reasoning, i.e. reasoning that could be attributed to rational agents. First, I drew a distinction between logical laws and logical norms, then I discussed the principles and put them to the test against the Preface and the Lottery paradoxes. I argued that the logical norms formulated as *reasons* for belief dealt better with the epistemic paradoxes than their strictly prescriptive counterparts. However, I pointed out that it has been argued that reasons don't have sufficient constraining power on our beliefs. I argued that none of the bridge principles is able to both (i) provide a plausible way to connect valid arguments and informal reasoning, and (ii) provide requirements for rational belief. Then, we saw that logical norms fall under the category of *prima facie* norms, which still retain prescriptive power (in the broad sense of 'prescriptive' mentioned in the chapter).

To sum up, in my thesis, I have argued that the normativity of logic is of a weak sort, which can be assimilated to the normative force of reasons, provided that we are ready to include them within our set of deontic operators. In the fourth chapter, I suggested two interpretations of the thesis dubbed *(ILR)*: (i) Logical entailment imposes obligations on what we should believe in order for us to think rationally, or (ii) Logical entailment gives us recommendations for what we should believe in order for us to think rationally.

In light of the final chapter, we can see that the two intepretations of (ILR) have different fates: (i) is ruled out by the the epistemic paradoxes examples, whereas (ii) remains plausible after the Preface and the Lottery tests. Was Harman right or wrong, then? It really depends on what we take (ILR) to mean. If, as I suspect, (ILR) amounts to (i), then Harman was right.

Bibliography

Adams (1975). *The Logic of Conditionals: An Application of Probability to Deductive Logic*. Dordrecht: Reidel.

Alchourron, G. and Makinson (1985). On the logic of theory change: Partial meet contraction and revision function. *Journal of Symbolic Logic*, 50(2):510–530.

Bealer (1998). Intuition and the autonomy of philosophy. In DePaul and Ramsey, editors, *Rethinking Intution: the Psychology of Intuition and its Role in Philosophical Inquiry*. Lanham: Rowman and Littlefield.

Boghossian (1997). Analyticity. In Hale and Wright, editors, *A Companion to the Philosophy of Language*, pages 331–368. Oxford: Blackwell.

Boghossian (2000). Knowledge of logic. In Boghossian and Peacocke, editors, *New Essays on the A Priori*. Oxford: Oxford University Press.

Bolzano (1972). *Theory of Science*. Oxford: Blackwell.

BonJour (1998). *In Defence of Pure Reason*. Cambridge: Cambridge University Press.

Brandom (2000). *Articulating Reason. An Introduction to Inferentialism*. Cambridge, Mass.: Harvard University Press.

Broome (1999). Normative requirements. *Ratio*, 12:398–419.

Broome (2013). *Rationality Through Reasoning*. Hoboken, NJ: Wiley-Blackwell.

Bueno (2010). Is logic a priori? *The Harvard Review of Philosophy*, 17:105–117.

Burge (1993). Content preservation. *The Philosophical Review*, 102(4):457–488.

Byrne and Johnson-Laird (2009). 'If' and the problems of conditional reasoning. *Trends in Cognitive Sciences*, 13,7:282–287.

Chellas (1974). Conditional obligation. In Stenlund, editor, *Logical Theory and Semantic Analysis*. Dordrecht: Reidel.

Chellas (1980). *Modal Logic: An Introduction*. Cambridge: Cambridge University Press.

Cherniak (1986). *Minimal Rationality*. Cambridge, Mass.: MIT Press.

Christensen (2004). *Putting Logic in its Place: Formal Constraints on Rational Belief*. Oxford: Oxford University Press.

Cosmides (1989). The logic of social exchange: Has natural selection shaped how humans reason? studies with the wason selection task. *Cognition*, 31:187–276.

Cosmides (2005). Detecting cheaters. *Trends in Cognitive Sciences*, 9(11):505–506.

Dretske (2000). Entitlement: Epistemic rights without epistemic duties? *Philosophy and Phenomenological Research*, 60(3):591–606.

Duca (2009). Rationality and the wason selection task: a logical account. *Psyche*, 15(1):109–131.

Dummett (1976). Is logic empirical? In Dummett, editor, *Truth and Other Enigmas*, pages 269–289. Cambridge, Mass.: Harvard University Press.

Dummett (1991). *Logical Basis of Metaphysics*. London: Duckworth.

Edgington (2008). Conditionals. In Zalta, editor, *The Stanford Encyclopedia of Philosophy*, volume URL=<http://plato.stanford.edu/archives/win2008/entries/conditionals/>. Winter 2008 edition.

Elqayam and Evans (2011). Subtracting "ought" from "is": Descriptivism versus normativism in the study of human thinking. *Behavioral and Brain Sciences*, 34:233–290.

Etchemendy (1990). *The Concept of Logical Consequence*. Cambridge, Mass.: Harvard University Press.

Evans and Lynch (1973). Matching bias in the selection task. *British Journal of Psychology*, 64(3):391–397.

Evans and Over (1997). Are people rational? yes, no and sometimes. *Psychologist*, 10(9):403–406.

Field (1998). Epistemological nonfactualism and the aprioricity of logic. *Philosophical Studies*, 92:1–24.

Field (2005). Recent debates about the a priori. In Gendler and Hawthorne, editors, *Oxford Studies in Epistemology*, pages 69–88. Oxford: Oxford University Press.

Field (2009). What is the normative role of logic? *Aristotelian Society Supplementary Volume*, 83(1):251–268.

Field (forthcoming). What is logical validity? In Caret and Hjortland, editors, *Foundations of Logical Consequence*. Oxford: Oxford University Press.

Foley (1993). *Working without a Net: A Study of Egocentric Epistemology*. New York: Oxford University Press.

Frankish (2009). Partial belief and flat-out belief. In Huber and Schmidt-Petri, editors, *Degrees of Belief*.

Frege (2013). *Basic Laws of Arithmetic*. Oxford: Oxford University Press.

Gentzen (1969). Investigations concerning logical deduction. In Szabo, editor, *The Collected Papers og Gerhard Gentzen*.

Glüer and Wikforss (2010). The normativity of meaning and content. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*, volume URL = <http://plato.stanford.edu/archives/win2010/entries/meaning-normativity/>. Winter 2010 edition.

Griggs and Cox (1982). The elusive thematic-materials effect in wason's selection task. *British Journal of Psychology*, 73:407–420.

Hanna (2006). *Rationality and Logic*. Cambridge, Mass.: MIT Press.

Harman (1982). Logic, reasoning, and logical form. In Simon and Scholes, editors, *Language, Mind, and Brain*. Hillsdale, NJ: Erlbaum.

Harman (1986). *Change in View: Principles of Reasoning*. Cambridge, Mass.: Harvard University Press.

Hintikka (1962). *Knowledge and Belief*. Ithaca, NY: Cornell University Press.

Hintikka (1970). Information, deduction and the a priori. *Nous*, 4(2):135–152.

Johnson-Laird (1983). *Mental Models: Toward a Cognitive Science of Language, Inference and Consciousness*. Cambridge, Mass.: Harvard University Press.

Johnson-Laird (2006). *How we reason*. Oxford: Oxford University Press.

Johnson-Laird (2008). Mental models and deductive reasoning. In Adler and Rips, editors, *Reasoning: Studies of Human Inference and Its Foundations*, pages 208–222. Cambridge: Cambridge University Press.

Kahneman (2011). *Thinking, Fast and Slow*. New York: Farrar, Starus and Giroux.

Kahneman and Tversky (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185:Republished in Kahneman 2011.

Kahneman and Tversky (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4):293–315.

Kaplan (1978). On the logic of demonstratives. *Journal of Philosophical Logic*, VIII:81–98.

Kelly (2003). Epistemic rationality as instrumental rationality. *Philosophy and Phenomenological Research*, 66(3):612–640.

Kolodny (2005). Why be rational? *Mind*, 114(455):509–563.

Kolodny and MacFarlane (2010). Ifs and oughts. *Journal of Philosophy*, 107(3):115–143.

Kyburg (1961). *Probability and the Logic of Rational Belief*. Middletown: Wesleyan University Press.

Lakatos (1976). *Proofs and Refutations*. Cambridge: Cambridge University Press.

Lewis (1973). *Counterfactuals*. Oxford: Blackwell.

MacFarlane (2000). *What Does It Mean To Say That Logic Is Formal?* PhD thesis, University of Pittsburgh.

MacFarlane (2002). Frege, Kant, and the logic in logicism. *Philosophical Review*, 111(1):25–65.

MacFarlane (2004). In what sense (if any) is logic normative for thought. Presented at the APA Conference 2004.

Makinson (1965). Paradox of the preface. *Analysis*, 25:205–207.

Nozick (1993). *The Nature of Rationality*. Princeton: Princeton University Press.

Oaksford and Chater (2007). *Bayesian Rationality*. Oxford: Oxford University Press.

Oaksford, Chater, and Hahn (2008). Human reasoning and argumentation: The probabilistic approach. In Adler and Rips, editors, *Reasoning: Studies of Human Inference and Its Foundations*, pages 383–413. Cambridge: Cambridge University Press.

O'Brien (1995). Finding logic in human reasoning requires looking in the right places. In Newstead and Evans, editors, *Perspectives on Thinking and Reasoning: Essays in Honor of Peter Wason*, chapter Finding Logic in Human Reasoning Requires Looking in the Right Places, pages 189–216. Hove: Erlbaum Associates.

Peirce (1932). *Collected Papers of Charles Sanders Peirce*, volume 1-2. Cambridge, Mass.: Belknap Press.

Peirce (1935). Minute logic. In Hartsthorne and Weiss, editors, *Collected Papers of C.S. Peirce, Vol.2*.

Piaget (1972a). *Essai de la logique opératoire*. Paris: Dunod.

Piaget (1972b). *The psychology of the child*. New York: Basic Books.

Pollock (1987). Epistemic norms. *Synthèse*, 71(1):61–95.

Popper (1946). Why are the calculuses of logic and arithmetic applicable to reality? *Aristotelian Society Supplementary Volume XX*.

Popper (1959). *The Logic of Scientific Discovery*. London: Routledge.

Priest (2006). *Doubt truth to be a liar*. Oxford: Clarendon Press.

Prior (1960). The runabout inference ticket. *Analysis*, 21:38–39.

Putnam (1979). The logic of quantum mechanics. In Putnam, editor, *Mathematics, Matter and Method (2nd edition)*, pages 174–197. Cambridge: Cambridge University Press.

Quine (1951). Two dogmas of empiricism. *Philosophical Review*, 60:20–43.

Railton (2000). A priori rules: Wittgenstein on the normativity of logic. In Boghossian and Peacocke, editors, *New Essays on the A Priori*, pages 170–196. Oxford: Oxford University Press.

Read (1995). *Thinking About Logic: An Introduction to the Philosophy of Logic*. Oxford: Oxford University Press.

Read (forthcoming). Proof-theoretic validity. In Caret and Hjortland, editors, *Foundations of Logical Consequence*. Oxford: Oxford University Press.

Rescher and Manor (1971). On inference from inconsistent premises. *Theory and Decision*, 1:179–217.

Restall (2004). Logical pluralism and the preservation of warrant. In et al., R., editor, *Logic, Epistemology and the Unity of Science*. Dordrecht: Kluwer.

Restall (2005). Multiple conclusions. In Glymour, Wang, W., editor, *Logic, Methodology and Philosophy of Science: Proceedings of the Twelfth International Congress*.

Rips (1994). *The Psychology of Proof: Deductive Reasoning in Human Thinking*. Cambridge, Mass.: Harvard University Press.

Rips (2008). Logical approaches to human deductive reasoning. In Adler and Rips, editors, *Reasoning: Studies of Human Inference and Its Foundations*, pages 187–205. Cambridge: Cambridge University Press.

Shapiro, L. (2011). Deflating logical consequence. *The Philosophical Quarterly*, 61(243):320–342.

Stalnaker (1970). Probability and conditionals. *Philosophy of Science*, 37(1):64–80.

Stanovich (2008). Individual differences in reasoning and the algorithmic/intentional level distinction in cognitive science. In Adler and Rips, editors, *Reasoning: Studies if Human Inference and Its Foundations*, pages 414–436. Cambridge: Cambridge University Press.

Stein (1997). *Without Good Reason: The Rationality Debate in Philosophy and Cognitive Science*. Oxford: Clarendon Press.

Stenning (2002). *Seeing Reason*. Oxford: Oxford University Press.

Tarski (2002). On the concept of following logically (translated from German and Polish by Magda Stroinska and David Hitchcock T.). *History and Philosophy of Logic*, 23 (3):155–196.

Thomson (2008). *Normativity*. Chicago, IL: Open Court Press.

Wason (1966). Reasoning. In Foss, editor, *New Horizons in Psychology*, pages 135–151. London: Penguin.

Wason (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20:273–281.

Wason and Johnson-Laird (1972). *Psychology of reasoning: Structure and content*. Cambridge, Mass.: Harvard University Press.

Wason and Shapiro (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, 23:63–71.

Williamson (2007). *The Philosophy of Philosophy*. Oxford: Blackwell.

Williamson (2013). How deep is the distinction between a priori and a posteriori knowledge? In Casullo and Thurow, editors, *The A Priori in Philosophy*, pages 291–312. Oxford: Oxford University Press.

Wright (2004). Warrant for nothing (and foundations for free)? *Aristotelian Society Supplementary Volume*, 78(1):167–212.