Logic, normativity and the *a priori*

MSc Thesis (*Afstudeerscriptie*)

written by

Theodora Achourioti
(born October 9, 1975 in Athens, Greece)

under the supervision of **Prof Dr Michiel van Lambalgen**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam*.

Date of the public defense:	**Members of the Thesis Committee:**
September 12, 2007	Prof Dr Peter van Emde Boas
	Prof Dr Dick de Jongh
	Prof Dr Martin Stokhof

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

In memory of Miltos

Contents

Preface

The topic of this thesis developed from an interest that dates back to my time as a philosophy student in Athens. At the time, thinking about the problem of realism drove me to Kant's *Critique of Pure Reason*. During my years in Amsterdam, first as a student of argumentation and then as a student of logic, I was particularly drawn to the question of how normative claims about reasoning can be justified. I decided to take this as the topic for my master of logic thesis in the conviction that Kant's transcendental view of normativity as developed in the first *Critique* is highly relevant here. This work is an attempt to justify this conviction.

I would like to thank my supervisor Michiel van Lambalgen, whose feedback has been essential to the writing of this work. What I am most grateful for is learning to appreciate the connection between technical results in logic and traditional philosophical considerations. Working as a student assistant on his courses has also provided me with a constant source of inspiration besides contributing to the financial support for my research. Chapters 2 and 6 owe much to discussions with Dick de Jongh, although responsibility for the claims made there is entirely mine. I am indebted to Peter Houtlosser for ideas motivating Chapter 3, in particular the distinction between internal and external normativity. Last but not least, this is a fitting moment to thank my supervisor from my Athens days, Stelios Virvidakis; in many aspects this is merely a continuation of the ideas that his teaching inspired.

CHAPTER 1

Introduction: is logic *a priori*?

1. High normative expectations associated with logic

Until quite recently, logic has been closely associated with rational, normatively correct reasoning. One system of formal logic, classical first order logic, was privileged to give the rules for rationally justified inferences, and to classify some argument patterns as fallacies. Educating a person in rational thinking was synonymous to teaching that person how to reason in classical logic. This attitude dates back to the founding fathers of logic. In fact, one of Frege's arguments against psychologism, the reduction of logical norms to empirical laws of thinking, was precisely that such a psychological view of norms makes it impossible to arbitrate between good and bad reasoning.

One finds this view of the role of logic forcefully represented in the textbooks. Copi's *Symbolic Logic* [**9**], very influential in its time, provides a quote from C.S. Peirce,

> It will generally be conceded that its [Logic's] central problem is the classification of arguments, so that all those that are bad are thrown into one division, and those which are good into another [**9**, p. 1].

before making Peirce's aim its own:

> The study of logic, then, is the study of the methods and principles used in distinguishing correct (good) from incorrect (bad) arguments (*ibid.*).

Further examples are easy to find. In his *Introduction to Logic* [**46**], Suppes explains to his readers how they should think of the contents of his book in the following words:

> In modern times logic has become a deep and broad subject. We shall initially concentrate on that portion of it which is concerned with the theory of correct reasoning, which is also called the theory of logical inference, the theory of proof or the theory of deduction. The principles of logical inference are universally applied in every branch of systematic knowledge. It is often said that the most critical test of any scientific theory is its usefulness and accuracy in predicting phenomena before the phenomena are observed. Any such prediction must involve application of the principles of logical inference [**46**, p. XV] (emphasis added).

And somewhat later he adds:

> Indeed it is not too much to claim that the theory of inference is pertinent to every serious human deliberation (*ibid.*, p. XVI).

On such views, logic is characteristically assigned the task of providing the tools that will yield secure scientific knowledge. These tools are typically identified with patterns of inference that are deemed necessary in that they guarantee certainty in and of themselves. Logic is also universal in that its normative principles permeate all reasoning that is correct.

This follows from a particular understanding of necessity as involving detaching context and domain considerations from what is essential to thought.

These ideas are also echoed in a much more recent textbook, *The Language of First-Order Logic*, by Barwise and Etchemendy [**3**]. In an introductory section entitled 'The special role of logic in rational inquiry', they write:

> What do the fields of astronomy, economics, law, mathematics, physics, and sociology have in common? [...] These fields all presuppose an underlying acceptance of basic principles of logic. For that matter, all rational inquiry depends on logic, on the ability of people to reason correctly most of the time. [...] While people may not all agree on a whole lot, they do seem to be able to agree on what constitutes a legitimate conclusion from given premises. Acceptance of these commonly held principles of rationality is what differentiates rational inquiry from other forms of human activity. Just what are the principles of rationality that underwrite these disciplines? And what are the techniques by which we can distinguish valid argumentation from invalid argumentation? More basically, what is it that makes one claim follow from accepted premises, while some other claim does not? Many answers to these questions have been explored. One suggestion that still has its adherents is that the laws of logic are a matter of convention. If this is so, we could presumably decide to change the conventions and so adopt different principles of logic, the way we can decide which side of the road we drive on. But there is an overwhelming intuition that the laws of logic are somehow more irrefutable than the laws of the land, or even the laws of physics [**3**, p. 1].

Note the implied tension between the conventionality and the irrefutability of logic. The irrefutability itself is never questioned.

2. Assumptions underlying the traditional view of logic

For a number of reasons that will be examined in greater detail below, the status of logic has somewhat diminished and few people today would subscribe to the high expectations that Frege envisaged for logic. This section outlines some of the customary assumptions that underlie the close association of rationality and classical logic.

2.1. Logical rules are universal norms. As several of the above quotations have shown, it is customary to assume that logic is fundamental to all other sciences. It is the supreme court to which differences of opinion can be referred. The universality of logic means that there can be only one (maximal) set of valid inferences, hence also that there is an absolute distinction between fallacious and non-fallacious inferences. Many people believe that a logic worth its name must have this universal character. If logic is not universal, it cannot be used to resolve disputes, for instance because the validity of inferences becomes itself a topic of argument, thus leading to a regress. A common justification of the universality of classical logic runs as follows. As will be seen in a moment, logical laws are considered to be valid by virtue of their form. This means that they are schematic, with variables for non-logical content. Since these variables can be replaced with arbitrary content, logic is content-independent and hence universally applicable.

2.2. Logical rules are valid in view of their form. This is usually taken as the defining characteristic of logic, but the concepts 'valid' and 'form' merit closer scrutiny. The naive idea can be put as follows. Every argument in natural language can be translated into a formal language in which a distinction is made between 'logical constants' and variable elements; the corresponding formulas in the formal language are then taken to be the *form* of the argument. To quote a random example:

> In studying methods of reasoning, logic is interested in the form rather than the content of the argument. For example, consider the two deductions:
> (1) All men are mortal. Socrates is a man. Hence Socrates is mortal.
> (2) All rabbits like carrots. Sebastian is a rabbit. Hence, Sebastian likes carrots.
> Both have the same form: All A are B. S is an A. Hence S is a B. The truth or falsity of the particular premisses and conclusions is of no concern to the logician. He wants to know only whether the truth of the premises imply the truth of the conclusion [**29**, p. 1].

Let us call this conception of logical form 'syntactic', because it only refers to the 'grammar' of the formal language. To see that this conception is non-trivial, note that assigning deduction 1 and 2 the same logical form is a considerable idealisation. For, whereas it is undoubtedly universally true that all men are mortal, the major premise of the second deduction is not obviously universally true, since new-born rabbits are an exception. So, one might well argue that the universal premises in 1 and 2 are both true, but that the universal quantifier 'all' means different things in examples 1 and 2, a difference which is obliterated by assigning the same syntactic form.

Using this notion of form, validity can then be defined, following Bolzano and Tarski, as: every substitution for the variable elements that makes the premises true, also makes the conclusion true.[1]

2.3. Logic is concerned with the products of thought, not with thought processes. Logicians in the Fregean tradition claim one need not be interested in how the reasoning subject performs a complex inference such as a syllogism, or indeed acquires simple inferences such as *modus ponens*. The logician is only concerned with what can be termed the extensional aspect of validity: the set of pairs <premises, conclusion> such that the conclusion validly follows from the premises. Copi is explicit on this point:

> The logician is not concerned with the process of inference, but with the propositions which constitute the initial and end points of this process, and the relationships between them. [...] Reasoning is that special kind of thinking called inferring, in which conclusions are drawn from premises. As thinking, however, it is not the special province of logic, but part of the psychologist's subject matter. Psychologists who examine the reasoning process find it to be extremely complex and highly emotional, consisting of awkward trial and error procedures illuminated by sudden /and sometimes apparently irrelevant /flashes of insight. These factors are all of importance to psychology. But the logician is not interested in the actual process of reasoning. He is concerned with the correctness of the completed process. His question is always: does the conclusion reached follow from the premises used or assumed? If the premises constitute grounds or good evidence for the conclusion, so that asserting the premises to be true warrants asserting the conclusion to be true also, then the reasoning is correct. Otherwise it is incorrect. The logician's methods and techniques have been developed primarily for the purpose of making this distinction clear. The logician is interested in all reasoning, regardless of its subject matter, but only from this special point of view [**9**, p. 2].

The message from this passage is that there is no systematic relation between the validities of logic and how people come to grasp these validities.

[1] Here it is of course assumed that the notion of truth is understood.

Interestingly, this is one area where many psychologists would agree with Copi: to explain reasoning in classical logic one cannot use this logic itself; various heuristics are necessary, for instance, those described by Johnson-Laird's 'mental models' theory [**25**]. Briefly, Johnson-Laird's theory can be viewed as opposing the theory of 'mental logic' which holds that logical reasoning proceeds by and large by application of natural deduction rules, and so remains within what classical logic has to offer. Johnson-Laird believes that people do not reason by applying rules but instead by constructing models of the situation described by the premises. This procedure is heuristic, however, because it may be beyond human capacities to construct all models corresponding to a given set of premises, and hence arguments may be declared valid which in fact are invalid.

A related point is that logic has nothing to say about the processes involved in getting from a natural language argument to a formalised version. Here is Copi again:

> the communication of propositions and arguments requires the use of language, and this complicates our problem. Arguments formulated in English or any other natural language are often difficult to appraise because of the vague and equivocal nature of the words in which they are expressed, the ambiguity of their construction, the misleading idioms they may contain, and their pleasing but deceptive metaphorical style. The resolution of these difficulties is not the central problem for the logician, however, for even when they are resolved, the problem of deciding the validity or invalidity of the argument remains (Copi [**9**, p. 7]).

Note that the phrase 'this complicates our problem', is strongly suggestive of the old positivist dream of a purified formal language replacing murky natural language. In other words, logic is viewed as obscured by natural language, which has to be stripped of many of its features before logical form becomes visible. This seems to suggest that logic hardly plays a role in natural language comprehension, and also that logic is a separate domain. A slightly more subtle version is found already in Russell:

> Some kind of knowledge of logical forms, though with most people it is not explicit, is involved in all understanding of discourse. It is the business of philosophical logic to extract this knowledge from its concrete integuments, and to render it explicit and pure ([**38**]; cited after Sainsbury [**39**, p. 1].

Taken together, these three assumptions have the effect of defining logic as a set of formally defined inference patterns claiming universal validity, but with little or no relation to actual human reasoning.

3. The 'normative status' of logic

Several forces have conspired to blur this picture. First, alternative views of rationality have come to the fore, both from within and without logic. To start with the latter, it has for example been observed that real-life reasoning is mostly concerned with decision under uncertainty, and that therefore classical logic fails on two counts: it does not deal with actions, and it cannot handle uncertainty. New standards of rationality have been proposed, for instance inductive reasoning in the form of probability theory. Within logic itself one has seen the development of various logics for applications in artificial intelligence, and equally importantly a sustained critique of classical logic by various forms of constructivism. To give just one example of the latter, Prawitz [**34**] has criticised the classical notion of validity given above as being a definition of validity instead of a consequence of validity. He argues that the Bolzano definition is useless for practice, where one would want to infer the truth of a conclusion of a valid argument from the truth of the premises, without having the capacity to survey all concrete instances of the particular argument form. Prawitz argues for the need of a notion of validity with lower epistemic requirements. The definition of validity that Prawitz proposes entails the Bolzano – Tarski criterion of validity without being epistemically equivalent to it, since validity is defined independently of truth.

The question raised by these developments is obviously that of justification; for instance, if the formal character of logic confers universality upon it, how is it possible that other logics exist and even find applications in places where classical logic fails?

Second, psychological investigations have opened up a considerable gap between what (classical) logic says is rational, and the inferences that people actually endorse. At first, psychologists tended to interpret these results as showing that human beings are irrational. In more recent times, however, the most prominent interpretation has been that irrational human beings would not have survived and that therefore the supposed criteria of rationality must be wrong. Alternative models of rationality have therefore been proposed, some, for instance, inspired by probability theory, and experimental evidence has been gathered supporting the claim that humans are rational according to *these* rational standards.

Third, the expectations people have had of logic: that it has normative force, that it will actually help one to determine good and bad arguments, have been disappointed, at least in the perception of many whose daily business it is to teach argumentation and 'critical thinking'. Copi could once write:

> ...the study of logic, especially symbolic logic, like the study of any other exact science, will tend to increase one's proficiency in reasoning [**9**].

but the current received wisdom points in a different direction. A great deal of criticism is the result of disappointment as to what logic can offer in the realm of real arguments, not only when these are everyday disagreements but even in theoretical contexts. Lecturers very often describe the difficulties they face when confronted with having to teach formal methods of argument in the classroom, while at the same time convincing the students of the applicability of those methods outside the scope of mathematical inquiry. Fisher [**14**] describes this frustration in the preface of a book in which he sets out to discuss argumentation analysis with respect to particular examples that give the student some idea about how to perform an analysis in a systematic logical manner.

> Like many others I hoped that teaching logic would help my students to argue better and more logically. Like many others, I was disappointed. Students who were well able to master the techniques of logic seemed to find that these were of very little help in handling real arguments. The tools of classical logic – formalisation, truth-tables, Venn diagrams, semantic tableaux, etc.– just didn't seem to apply in any straightforward way to the reasoning which students had to read in courses other than logic [**14**, p. vii].

In fact, an entire field of study, argumentation theory, has emerged mainly in order to cover areas unexploited by, or, as is often assumed, unreachable to, logically-oriented investigations [**51**]. The opposition to the disadvantages that were brought up by a 'logical approach' in the study of argument, has been the essential point of reference in defining this new field of study. This disadvantage is actually viewed as a necessary consequence of what logic pretended to be. In real argumentation, there is no such thing as absolute normativity (i.e. absolute standards of evaluation) and the validity of arguments is relative to the circumstances at hand.

4. Outline of the thesis

Section 1 outlined a common conception of logic according to which its laws are universal and necessary. Following Kant's usage in the *Critique of pure reason*, B4[2] such laws are also called *a priori*. Section 2 cast doubt on the conclusion that logical laws are *a priori* in this sense. However, if one revisits Kant, one sees that the first meaning of *a priori* introduced is 'independent of all experience and even of all impression of the senses'. Kant believed that universality and necessity are hallmarks of cognitions which are *a priori* in the latter sense. But for our present discussion it is useful to keep the two senses apart. For Kant, the two senses were connected due to the special meaning he gave to the word 'necessary', which for him means: necessary in view of the way our cognition is constituted. Kant believed that there are general constitutive rules valid for all of cognition, and therefore necessary for him also meant: universal. However, if one adopts a more local view of cognition, these concepts become separated. It is this possibility that will be investigated in what follows, i.e. logics as necessary to both the constitution and performance of cognitive tasks. Logic is then both *a priori* in the sense of being independent from all experience and in the sense of being necessary, because without logic no cognitive task comes to exist. The claim to universality must be dropped however, because different cognitive tasks may require different logics.

The aim of this thesis is to explore how an informed and revised notion of normativity reclaims the *a priori* character of logic. Chapter 2 explores the notion of fallacy and draws upon experimental results to show that there is a gap between traditional standards of rationality and how people actually reason. Furthermore, it argues that what people do in these experiments is nonetheless reasonable. This leads to a redefinition of the notion of fallacy. In the final section, the chapter introduces an idea that is fundamental to the thesis, namely that in reasoning one must go 'beyond the information given'. In an attempt to add further precision to the concept of rationality, Chapter 3 distinguishes two kinds of normativity: external normativity, which relies on norms supposed to be already given, contrasted with internal normativity, which considers norms that are in a sense inherent in cognitive tasks. This topic is further developed in chapter 4, where internal normativity is explained using the concept of constitutive norms as defined by Kant in the first *Critique*. Constitutive norms are tied up with a fundamental feature of cognition, namely that a process of synthesis is necessary to produce coherent cognitions at all. This is the topic of Chapter 5. In Chapter 6, logic again comes to the fore. It is argued that logic actually embodies two kinds of norms, constitutive and regulative norms. The first attempt to make this idea more precise is via the proof-theoretic semantics pioneered by Dummett and Prawitz. Very roughly speaking, one may identify the introduction rules in natural deduction system with the constitutive norms, whereas the elimination rules fall under the rubric of the regulative norms. In the end, these ideas are found wanting, however, and the next chapter proposes a much more general definition. In Chapter 8, the focus is on reasoning tasks. It is shown that, as a result of the preceding considerations, the relation between competence and performance in these tasks is vastly more complex than entertained in current psychology of reasoning. Finally, Chapter 9 returns to the original question: in what sense is logic *a priori*?

[2]Cited after the English translation by Paul Guyer [**27**].

CHAPTER 2

Fallacies lost, fallacies regained?

The previous chapter observed that logic was traditionally conceived as giving an absolute distinction between valid and fallacious argument. The present chapter queries whether such an absolute distinction exists.[1] This will be seen to tie in with another issue raised in chapter 1, namely the prevalent opinion among argumentation theorists, that logic has very little, if anything, to offer to the study of argumentation. The chapter concludes with some more technical remarks concerning the notions of logical form and interpretation which will help us discuss in what sense logic is normative in later chapters.

1. Fallacies and argumentation

Fallacies are central in the study of argumentation; it has been a perpetual challenge for argumentation theorists to distinguish types of fallacies and to formulate unambiguous and systematic definitions that make it possible for fallacious arguments to be identified. For a long time, classical logic provided the hope that such definitions could be formulated in terms of argument patterns that can be found in natural language. The argument patterns of denying the antecedent (DA) and affirming the consequent (AC) have been thus defined as the fallacious counterparts of the unquestionably valid modus ponens (MP) and modus tollens (MT). Although the use and relevance of classical logic has been questioned time and again in the study of argumentation, the fallacies of DA and AC have remained as the bare minimum of logic to be regarded relevant to argumentation. Such has been the connection of these fallacies with logic that argumentation textbooks still opt to call them the ‘logical fallacies’ to distinguish from the rest.

The question posed in this chapter is whether it is at all possible to define the fallacious inferences DA and AC in terms of argument patterns. The chapter begins with some preliminary remarks on the notion of logical form and the role it plays in establishing semantic interpretation (section two). The importance of interpretation will be highlighted in the third section, by reporting on some experimental results: it will be shown that validity is meaningful only relative to the logical form assigned to the argument at hand. The next sections explore the consequences of the observation that a logical form in which rules are defeasible plays a prominent role in natural language. The main consequence for argumentation is that the resulting notion of validity is context-dependent. This notion of validity will be made more precise in the fourth section. It will be illustrated (again by the help of experimental results) that the validity of the four inference types is relative to: a) the type of underlying reasoning (classical or defeasible) as well as b) the world-knowledge deemed relevant to the argumentation at hand. Defining the fallacious inferences in terms of argument patterns will then be no longer possible (sections five and six). It will eventually be claimed that formalising a more context-sensitive notion of fallacy can restore some of logic’s pertinence to the study of argumentation.

[1] Part of this chapter has been presented at OSSA 2007 and will appear in its Proceedings.

Below is depicted a set of four cards, of which you can see only the exposed face but not the hidden back. On each card, there is a number on one of its sides and a letter on the other.

Also below there is a rule which applies only to the four cards. Your task is to decide which if any of these four cards you *must* turn in order to decide if the rule is true. Don't turn unnecessary cards. Tick the cards you want to turn.

Rule: *If there is a vowel on one side, then there is an even number on the other side.*

Cards:

A	K	4	7

FIGURE 1. Wason's selection task

p	p, q	$p, \neg q$	$p, q, \neg q$	misc.
35%	45%	5%	7%	8%

TABLE 1. Typical scores in the selection task

2. Natural language and logical form

In order to talk about argument patterns and isolate the so-called logical fallacies, one has to distinguish the argument as it manifests itself in natural language from its formal representation. This is because the meaning of natural language expressions is far from being transparent. One cannot rely solely on the syntactic configurations and the occurrence of some key-words to guarantee a common ground understanding of the logical form of natural language expressions. What the formal representation does, then, is to fix an interpretation that is accurate and precise enough to determine the standards against which an argument can be evaluated. In what follows, formal representation is understood as meaning interpretation in this sense.

As explained in the previous chapter, classical logic has long been regarded as the arbiter of thought, and this bias has given rise to some dramatic twists and turns in the history of the study of reasoning. At first, experiments like the famous Wason's selection task [**53**] led theorists to think very poorly of the logical capacities of the general population, which gave rise to considerable scepticism concerning the general standards of rationality. In the selection task, subjects are given the following instructions (figure 1).

These experiments basically tested whether the subjects were able to solve the given task by means of classical logic. Not providing the classically right answer (i.e. A and 7) was categorically marked as a sign that the subjects were not able to reason logically. If one formulates the rule

> *If there is a vowel on one side, then there is an even number on the other side.*

as an implication $p \to q$, then the observed pattern of results is typically given as in table 1. The vast majority of subjects would, thus, be irrational.

Soon, vehement opposition to these hasty conclusions was put forward. All that these experiments have managed to achieve is to show in the starkest manner how much distance

Below is depicted a set of four cards, of which you can see only the exposed face but not the hidden back. On each card, there is a 8 or 3 on one of its sides and a U or I on the other.

Also below there are two rules which apply only to the four cards. It is given that exactly one rule is true. Your task is to decide which if any of these four cards you *must* turn in order to decide which rule is true. Don't turn unnecessary cards.

(1) *if there is a U on one side, then there is an 8 on the other side*
(2) *if there is an I on one side, then there is an 8 on the other side*

Cards:

U	I	8	3

Please, circle your choice out of the possible choices below:

FIGURE 2. Two–rule task

actually exists between classical logic and ordinary reasoning. Consequently, those interested in the more mundane task of comprehending how people actually reason were driven to repudiate classical logic, and the whole canon of formal logic fell along with it.

Formalising natural language is undoubtedly a difficult enterprise and the relevant problems are discussed in the literature at length. For the purposes of this chapter it is important to take notice of one particular aspect of the difficulties involved, which relates to the traditional notion of a logical form. For many years the conception of logical constants as the skeleton of natural language endured among logicians (the literature is extensive; three prominent references are: Copi [**9**]; Strawson [**45**]; Tarski [**47**]). In this view, conjunction, disjunction and the like are seen as such primitive concepts that language and thought cannot do without. Identifying the conjunctive meaning where it occurs was naturally seen as simply 'recovering' part of the logical meaning already inherent in the language, or alternatively, of its (underlying) logical form.

However, it is not at all clear, in fact, it is rather doubtful, whether such inherent meanings playing the role of logical constants can actually be identified in natural language. Take conjunction as an example: whereas in classical logic conjunction is a commutative connective, in ordinary use the meaning it takes is often a temporal, non-commutative one. At the same time, both notions are excessively used in natural language. In brief, there is no formal system to propose a one-to-one mapping from natural language expressions to logical constants such that it can be trusted to "recover" the logical form.

3. The importance of interpretation: the 'two rule task'

What the early experiments on reasoning tasks failed to appreciate is that the interpretation that an arguer has of linguistic discourse may systematically and consistently differ from that intended by the analyst. To illustrate this, consider an adaptation of the following reasoning problem, the 'two-rule task' (due to Stenning and van Lambalgen [**42**]). The task was replicated by the author on a group of 25 first-year students at *University College Utrecht* (UCU) in spring 2006. It was used as a kind of preliminary, warm-up exercise to precede the instruction of elementary propositional logic in an introductory course on argumentation. The task was also used later in the course as a means to illustrate material implication by explaining what the classically right answer would be. The experimental material is given in figure 2.

The students were provided here with choices like ‘no cards’, ‘only U’, ‘U & I’, ‘U or 3’, ‘U or I or 8 or 3’, ‘3 & (U or I)’, and so on: in total, 44 combinations of the letters, the numbers, conjunction and disjunction. In the classical understanding of the implication, the correct answer is to choose ‘only 3’. Surprisingly no student gave this answer. Instead 75% chose for ‘U or I or 8 or 3’ which meant that turning any card would do. The students were asked to provide arguments to justify their answers. Here is a characteristic justification of an answer opting for ‘U or I or 8’:

Subject 12: When you turn I and there is an 8 at the back, then the U-rule is false. When you turn the card with an 8 and there is an I on the back, then you prove the I-rule and when it has an U on the back you prove the I-rule wrong.

Faced with such an argument, the analyst has two possibilities: either to convict the student of committing a fallacy, or to inquire into the meanings that the students assigned to the conditional, as well as truth and falsity. This thesis is in favor of the second option, which allows for more logical forms than the one dictated by classical logic. Natural language contains several different notions of the conditional, and for anybody who has given a course on basic classical logic it is clear that material implication is not the most intuitive one. What is more, opting for a different notion of the conditional entails choosing notions of truth and falsity that can be different from those of classical logic. The linguist Fillenbaum observed that about half of his subjects interpreted the statement ‘$p \rightarrow q$’ is false as $p \rightarrow \neg q$ (Fillenbaum [**13**]). Applied to the present task this means that the U-rule is false iff U only goes with 3. In other words, this notion of falsity is much stronger than the classical notion, which allows both a ‘U-8’ and a ‘U-3’ combination. The vast majority of the students in this experiment seemed to adopt this stronger notion of falsity. Here is one more characteristic quote of a student opting for ‘U or I or 8 or 3’:

Subject 7: If you turn around any card, ‘I’ or ‘U’ or ‘8’ or ‘3’, you will find out the rule. If, for example, you turn around I and it has a 3 on the other side, you know that behind the U is an 8, that behind an 8 is a U, behind 3 an I and vice versa. Therefore, by turning around any of the four cards you can eliminate one of the two rules and prove that only one rule is correct. After all, an 8 cannot be on the back of a card of both an I and a U.

In this interpretation, the subjects’ most prominent answer that any card suffices is of course completely logical. So, this example shows that it may be impossible to simply read off a logical form from the linguistic expressions given. A fortiori it is impossible to charge the subjects with simply committing a fallacy.

Of course, the above should not be taken to imply that no evaluation of the subjects’ answers can be performed. In both cases mentioned the students’ answers directly followed from their interpretation. But there were also students whose answers were inconsistent with the arguments they provided . In such cases it does make sense to invoke the notion of a fallacy, albeit only after the interpretation of the subject has been determined. This last point will be further illustrated in the following section with the help of a variant of the so-called suppression task.

4. Validity and context-dependence

We now come to a much more radical deviation from the classical logical form, which will lead us to question the very possibility of defining valid and invalid argument patterns. A number of technical notions need first to be established. One common way to identify argument patterns in natural language is by means of the following procedure. First, some logical constants are identified; one then identifies atomic propositions, that is, the propositions that do not contain these logical constants, and replaces them with variables ‘p’, ‘q’ etc. The result is what is called an ‘argument pattern’ in natural language.

The next task is to define validity and non-validity for such an argument pattern. To do so, one needs a semantics and a definition of valid consequence. It is important to note that one cannot talk about validity in an absolute sense but only relative to a semantics and a definition of validity. If the logical constants are given by two-valued truth-tables, and if one furthermore defines valid consequence as ‘whenever the premises are true, so is the conclusion’, then there are two prominent valid inferences, MP and MT, and two prominent fallacies, DA and AC. Both the validity of MP and MT and the invalidity of DA and AC follow directly from the truth-table of material implication. As a consequence, these four argument patterns have the effect that adding context does not change the (in)valid status of the argument. In fact, it is this context-independence that allows one to separate the inferences made from background assumptions.

These seemingly trivial conceptualisations have been called into question both by theoretical developments in logic and by experiments on human reasoning. Below is a variant of a standard propositional inference task that investigates reasoning with conditionals in the presence of context. The task was given in autumn 2006 to three groups of students (also at UCU), with total population size 56; these groups were similar to the one that took part in the two-rule task, as were the motivation and the use of the task afterwards in the classroom.

The students were given argument patterns with either two or three premises and a putative conclusion, and were asked to judge whether the argument pattern was valid or invalid. Here is the first example of the list:

1) valid / invalid If she has an essay to write, she will study late in the library. If the library is open, she will study late in the library. She has an essay to write. She will study late in the library.

On the face of it, this is a MP inference with an extra conditional premise added. Since the argument pattern MP is invariant with respect to context, one would expect roughly the same amount of students to endorse this inference as those endorsing the two-premise version of MP, which is typically 95% of the population. This, however, did not turn out to be the case. In the population tested the percentage dropped to 60%, which is comparable to what was found in other studies.

This experimental paradigm was devised by Byrne [**6**] to argue against the so-called mental logic view of reasoning, which holds that reasoning consists in the application of rules such as MP. For even if a premise is added, the rule MP remains applicable. Byrne took her results to be support for the rival mental models theory, but it will be argued here that the implications of the experiment go much deeper and affect the very nature of reasoning. The full set of experimental materials is provided here for the convenience of the reader.

Essays and Libraries: A Reasoning Task. This is meant as a training exercise in reasoning. It is neither an intelligence test, nor an exam on propositional logic (the answers are not graded!).

(a) Determine whether the following argumentation is valid or invalid.

1) valid / invalid
If she has an essay to write, she will study late in the library.
If the library is open, she will study late in the library.
She has an essay to write.

She will study late in the library.

2) valid / invalid
If she has an essay to write, she will study late in the library.
If the library is open, she will study late in the library.
She will study late in the library.
She has an essay to write.

3) valid / invalid
If she has an essay to write, she will study late in the library.
If the library is open, she will study late in the library.
She does not have an essay to write.
She will not study late in the library.

4) valid / invalid
If she has an essay to write, she will study late in the library.
If the library is open, she will study late in the library.
She will not study late in the library.
She does not have an essay to write.

5) valid / invalid
If she has an essay to write, she will study late in the library.
If she has some textbooks to read, she will study late in the library.
She has an essay to write.
She will study late in the library.

6) valid / invalid
If she has an essay to write, she will study late in the library.
If she has some textbooks to read, she will study late in the library.
She will study late in the library.
She has an essay to write.

7) valid / invalid
If she has an essay to write, she will study late in the library.
If she has some textbooks to read, she will study late in the library.
She does not have an essay to write.
She will not study late in the library.

8) valid / invalid
If she has an essay to write, she will study late in the library.
If she has some textbooks to read, she will study late in the library.
She will not study late in the library.
She does not have an essay to write.

9) valid / invalid
If she has an essay to write, she will study late in the library.
She has an essay to write.

MP2	MPad	MT2	MTad	DA2	DAalt	AC2	ACalt
99%	60%	75%	43%	75%	10%	66%	24%

TABLE 2. Scores in the suppression task (UCU 2006)

She will study late in the library.

10) valid / invalid
If she has an essay to write, she will study late in the library.
She will study late in the library.
She has an essay to write.

11) valid / invalid
If she has an essay to write, she will study late in the library.
She does not have an essay to write.
She will not study late in the library.

12) valid / invalid
If she has an essay to write, she will study late in the library.
She will not study late in the library.
She does not have an essay to write.

Please, answer the following questions:
(b) If you have evaluated 4, 8, and 12 differently, explain why:

(c) If you have evaluated 2, 6, and 10 differently, explain why:

A few explanations are in order. The added premises are of two kinds: (a) If the library is open, she will study late in the library, and (b) If she has some textbooks to read, she will study late in the library. Premise (a) is called additional; intuitively speaking this type of premise introduces an additional condition for the conclusion to hold. Premise (b) is called alternative; it can be read as introducing another sufficient condition for the conclusion to hold. Additional premises affect the rate of endorsement of the inference MP and MT, whereas alternative premises affect the rate of endorsement of the traditional fallacies DA and AC. Table 2 presents some pertinent results. It should be read as follows. In the case of MT 75% of the tested population endorses the two-premise inference. This percentage drops to 43% in the presence of an additional premise. Similarly for the other inferences.

Before discussing the theoretical implications of these results, some of the characteristic arguments that the students provided for their answers are presented.

First, an answer to question (b) in the test:
Subject 19: As regards 8 and 12, since she does not go to the library apparently she has no essay to write. As regards 4, it could also be the case that she does not study late in the library because it is closed.

And as regards question (c):
Subject 32: I have evaluated 6 differently from 2 and 10 because 2 and 10 don't say anything about other possible reasons/assignments for the girl studying in the library.
Subject 55: In 10 there is no exception whereas in 2 and 6 there are other possibilities.

Strict adherence to classical logic would force us to conclude that the subjects who suppress MP and MT in the context of an additional premise are reasoning illogically, as it happens with those subjects who suppress DA and AC in the context of an alternative premise. However, in accordance with what was said earlier, such a conclusion would be rather insensitive to the logical forms that the subjects may give when interpreting the task they are given.

Examples such as these exploit the fact that conditionals in natural language are defeasible. In fact, most rules stated in natural language are defeasible.[2] The consequent follows from the antecedent only ceteris paribus. Here is a direct example of this defeasibility. Suppose we have a rule saying that if a patient has cystitis, she must be given penicillin (example taken from Johnson-Laird and Byrne [**24**]). If the patient with cystitis presents herself the doctor will give her penicillin, but he will no longer do so if he has the additional knowledge that she is allergic to penicillin. So what seems as a simple MP inference is in fact retracted. The so-called suppression of MP in the suppression task can be viewed as an instance of this phenomenon.

Stenning and van Lambalgen [**43**] claim closed world reasoning to be the appropriate logic for defeasible conditionals. According to closed world reasoning, what is not forced to be true can be assumed to be false. As a form of underlying reasoning, closed world reasoning can explain both the suppression of the classically valid inferences and MP and MT and the endorsement of the traditional fallacies of DA and AC. Applied to AC, closed world reasoning takes the following form. If $p \rightarrow q$ is the only known rule whose consequence is q, and if we know q to be the case, then we can conclude that p must have 'caused' q. The assumption behind this reasoning pattern is that the effect is generated by causes, so that if there is only one cause for a given effect, the cause must actually have occurred. For DA the argument runs as follows: since p is the only possible cause for the effect q, non-occurrence of p entails non-occurrence of q. This explanation also accounts for the so-called 'conditional perfection', that is, the tendency to read conditionals as bi-conditionals in some contexts (Geis and Zwicky [**17**]). Suppression of DA and AC is now easily explained, because the provision of the alternative premise $r \rightarrow q$ highlights a second possible cause r for the effect q.

Explaining the suppression of MP and MT now requires taking full account of the fact that natural language conditions are defeasible. A formal representation of a defeasible conditional can be given as 'p and nothing abnormal is the case implies q'; formally, $p \wedge \neg ab \rightarrow q$. With this representation the categorical premise p does not itself warrant the conclusion q. But here closed world reasoning comes to our rescue because if there is no positive information about exceptions, we may assume that they do not occur, giving us the second conjunct of the antecedent. Note that in this way we have justified the pattern MP for two premises on the basis of non-classical reasoning. This justification no longer works in the presence of the additional premise, for in that case the possibility that the library is closed highlights a possible exception to the rule.

It appears then that in closed world reasoning all four inference patterns become context-sensitive. This leads us to the question of whether there are general definitions of valid and invalid argument patterns that can replace the 'gang of four'.

[2] It is impossible to give an overview of the large literature on conditionals here. The reader is directed to Athanasiadou and Dirven [**1**] (cf. also [**2**]) for a corpus study.

5. Valid and invalid argument patterns

A natural language argument pattern (i.e. with intuitive semantics for logical operators) is valid (fallacious) if it is an instance of an (in)valid argument pattern in a logical form. In this section it will be inquired whether it is at all possible to define argument patterns that can be used with certainty for the familiar fallacies to be detected in natural language. First, a more detailed description will be attempted of how these argument patterns would look like for the two logical systems mentioned above, that is, classical logic and closed world reasoning. For reasons of brevity, only two of the four inferences will be treated here, MP and AC. The argument patterns for MP and AC in classical logic are defined as follows[3]:

(i) MP $=_{def}$ $p; p \rightarrow q/q$
(ii) AC $=_{def}$ $q; p \rightarrow q/p$

What is characteristic about these definitions is that they constitute sufficient information to decide the normative status of an argument. In other words, no additional information can (in)validate the inferences if it is indeed the case that p implies q. This phenomenon is known as monotonicity. It follows that no considerations need to be taken into account regarding the context in which such patterns occur; the definitions are complete as they stand. This notion of context can be made more precise by formulating it as theory T, that is, a set of sentences considered to be true in the language in which we are working (here propositional logic). The context-independence of the (in)validity of these argument patterns then means the following: if we have two theories S and T, and from S, p and $p \rightarrow q$ it follows that q, then this also holds with S replaced by T. And similarly for AC, that is, if S, q and $p \rightarrow q$ entail p.

By contrast, in non-monotonic logics, the above argument patterns would not do if they are to serve as absolute reference points for the (in)validity of arguments. At the very least and in accordance with what was observed in the previous section, a slot should be included in MP that would allow for what the subjects usually call the 'exception' to be incorporated, and which explains the suppression effect:

(iii) MP $=_{def}$ $p; p \wedge \neg ab \rightarrow q/q$

But this cannot yet serve as a pattern of a valid argument form, since the validity of the argument depends on whether such an exception actually holds. A second clause is then needed to ensure that no such exception can be proven. This can be intelligibly stated only relative to a particular context, that is, relative to a particular theory T. The enriched pattern would have to include one more clause:

(iv) MP $=_{def}$ $p; p \wedge \neg ab \rightarrow q; \neg ab/_T q$

Appealing to closed world reasoning, which featured prominently in the subjects' argumentation in the suppression task, the negation of the abnormality ab should be understood as follows. Suppose $\varphi_1 \rightarrow ab, \ldots, \varphi_n \rightarrow ab$ are all the clauses in T which have ab as a consequent. If no φ_i can be proven in T where $1 \leq i \leq n$, then we can conclude $\neg ab$.

Similar considerations hold for AC, only here the argument pattern (i) is valid in closed world reasoning. However, this can only be the case when nothing apart from what is indicated by the categorical premise p can be proven to bring about the consequent q. In other words, AC is valid only when alternative rules have been excluded, and this can again be determined only with respect to a particular theory T. The enriched argument pattern of AC would, therefore, have to include the following information:

[3]The forward slash separates premises and conclusions.

(v) AC $=_{def}$ $q; p \wedge \neg ab \rightarrow q/_T p$,

which should be read as follows: Suppose $\psi_1 \rightarrow q, \ldots, \psi_n \rightarrow q$ are all the clauses in T which have q as a consequent. If no ψ_i can be proven in T where $1 \leq i \leq n$, then we can conclude p.

However, these enriched definitions are no longer the type of argument patterns that can be used in order to detect whether a fallacy has occurred in natural language. Whereas in classical logic argument patterns are local and separated from the theory (that is, the general context), in non-monotonic logics an argument pattern for MP or AC turns out to be theory-dependent. This is because global considerations need to be taken into account in order for the validity or invalidity of the argument to be established. These considerations relate to exceptions in the case of MP and MT, and to alternative rules in the case of the traditional fallacies DA and AC.

6. Where did the fallacies go?

The latter observations are not meant to repudiate the legitimacy of singling out particular instances of DA or AC as cases of fallacious reasoning. In classical logic, DA and AC are never valid, and they are not always valid in closed world reasoning either. Therefore, it still makes sense to seek a definition that captures the fallacious forms of DA and AC, that is, a definition that tells us when the inference is wrong. However, the previous sections show that such a definition cannot be restricted to an argument pattern if an argument pattern is understood in the traditional intuitive sense.

The moral to be drawn here is that once a broader, i.e. semantically informed, notion of logical form is accepted together with the possibility of formal systems alternative to classical logic as plausible representations of how people think, the familiar landscape of absolutely valid and absolutely invalid inferences changes drastically. MP and MT are sometimes considered invalid, and DA and AC are sometimes considered valid. As a consequence, ‘DA’ (forward inference that negates the antecedent) and ‘AC’ (backward inference that asserts the consequent) turn out to be infelicitous terms for fallacious reasoning. In fact, they are not even suitable characterisations of the types of inferences they represent, since they do not cover, for instance, the argument patterns in natural language in which the constituent ‘ab’ does not occur overtly. Accounting for the fact that covert information can be instrumental in assigning a logical form to an argument and deciding its normative status, a generalised definition of the inferences would closely resemble to something like the following:

MP: Affirmation of the overt part of the antecedent in the overt rule.
MT: Denial of the overt consequent in the overt rule.
DA: Affirmation of the overt consequent of the overt rule.
AC: Denial of the overt antecedent of the overt rule.

Then it depends on the logic at hand to decide whether the inference is valid or not. The logic appealed to here in order to account for a number of the experimental results was a non-monotonic system, namely closed world reasoning. Non-monotonic systems make it possible to formalise defeasible rules, and as such they seem to be more appropriate for describing ordinary reasoning. Of course, different logics may pose additional challenges in defining fallacious argument patterns. In any case, the coarse grained contrast between monotonic and non-monotonic systems highlighted here is sufficient to support the general claims in this chapter, namely the impossibility of defining the particular inferences in terms of argument patterns that will unambiguously decide whether an argument is valid or not.

7. Logic is not insensitive to argumentation after all

The significance of a context-dependent notion of validity that one encounters in non-monotonic logics is fully appreciated when one realises how prominent defeasible reasoning is in ordinary argument. In fact, one could say that operating on defeasible inferences is what makes assumptions, rules and conclusions debatable, and gets argumentation started. Defeasible reasoning manifests itself every time we, as protagonists, retract an argument, or modify it; also whenever, as antagonists, we advance a counter-argument to rebut the reasons adduced by our opponent.[4]

The syntactic product-like representation entailed by classical logic has made it totally inappropriate for the modelling of ordinary argument. There is no room in such a representation to incorporate what might be disputable in the inference and may give rise to disagreement. As a consequence, there is also no room for the changes that a critical reaction might bring to the argument. This is precisely why a system like natural deduction for first order predicate logic breaks down when used to account for the inferential process that underlies argumentation. Take the additional premise in the suppression task as an example of a critical reaction. One wants to be able to explain how conjoining this premise with the initial rule gives rise to a conditional where the two antecedents are connected by conjunction instead of disjunction, as it would have to be the case according to classical logic. In other words, one wants to account for the fact that this premise is understood as an additional and not as an alternative one. On the face of such observations, it is no wonder that translating premises and conclusions into classically defined logical patterns has been considered irrelevant to argumentation as well as disconnected from the critical, dynamic processes there within.

At the same time, not all exceptions or alternative reasons are relevant and in need of being taken into account. Appealing to a high improbable cause or abnormal situation might be an entirely uncooperative move in the pragmatics of argumentative discourse. It is, therefore, very strange to be forced to call such “irrelevant” moves instances of valid argumentation. In classical logic, the truth of the antecedent does not follow from affirming the consequent because it is conceptually possible for p to be false. In other words, DA and AC are judged to be fallacious argument patterns on the presupposition that all circumstances can be checked in which the premises are true. However, closed world reasoning is based on exactly the opposite assumption, namely that not all possible circumstances are accessible; this is how incorporating new information can revise the interpretation by bringing to notice some situation that has not been deemed relevant at first.

In the present view, endorsement of DA or AC need no longer be seen as a logical mistake. A more fruitful perspective is to view these inferences as the arguer’s attempt to shape the common ground, that is, to establish the underlying reasoning in addition to shared world-knowledge. For instance, the arguer may attempt to elicit agreement for disregarding as irrelevant those exceptions that are not mentioned. This way one can also disentangle different types of critical reactions. Consider an AC inference based on a causal relation as an example. There is a difference between challenging the relation by adducing an alternative cause and by addressing the intensional meaning of the implication, that is, the sufficiency of the cause to bring about the effect (assuming that closed world reasoning

[4]Here I assume the pragma-dialectical terminology, according to which each of the two parties in a difference of opinion assumes the role(s) of ‘protagonist’ and/or ‘antagonist’ depending on the attitude towards the standpoint. See van Eemeren and Grootendorst [**51**].

holds). For example, it is one thing to deny that ‘she has an essay to write’ by arguing that there might be other reasons to have kept her late in the library, but quite another to do so by claiming that she is not the type of diligent student to work until late no matter how much work she has to do. In most cases, arguing for alternative reasons would entail the commitment of the speaker to the main conditional, so that ‘having an essay to write’ is accepted as a good enough reason to have kept her late in the library. This is because the main conditional is given, and usually not attacking given information is tantamount to including it in the common ground.

Closed world reasoning has some impact on argumentation, in particular on the notion of common ground. Traditionally, the common ground in argumentation is conceived of as consisting of shared beliefs and assumptions, which are to be used locally in argumentation: arguers appeal to these beliefs and assumptions individually to support their argumentation. In closed world reasoning, the common ground is used globally. The arguer about to use an AC inference $q; p \rightarrow q/p$ appeals to his opponent to concede that the common ground contains no rule of the form $r \rightarrow q$ for r different from p. Thus, the boundaries of the common ground become instrumental in enlarging the common ground itself. In this sense, the boundaries of common ground play an important strategic role in argumentation. Something similar holds for closed world reasoning as applied to exceptions: if the common ground gives us no reason to suppose an exception will occur, the non-occurrence of the exception can be added to the common ground.

It follows from the above that a more liberal view of interpretation alongside a semantically informed notion of logical form poses insurmountable obstacles to defining the so-called logical fallacies in terms of argument patterns. However, we do not repudiate the notion of fallacy altogether. Instead we have tried to inquire into what it is that the validity of the particular argument patterns is relative to. This led to a more precise meaning of the context-dependence of fallacies, which should be regarded as the first step towards formulating a more context-sensitive definition of all four inferences, MP, MT, DA and AC.

The validity of argument patterns has been found relative to both the underlying reasoning and -where this applies, e.g. in non-monotonic logics- to the assumptions of world-knowledge that pertain. At the same time it has been observed that defeasible inferences occupy a very prominent role in ordinary argument. These remarks are in themselves sufficient to explain why classical logic has been deemed inappropriate for the study of argumentation. However, formalising argumentation has been pursued here as the means to fix the interpretation, which makes it possible for the validity of the argumentation to be judged.

8. ‘Beyond the information given’

In this chapter, two reasoning tasks have been considered which lead to a variety of responses in conflict with classical logic. In some cases the answer patterns could be explained by adopting a different competence model, for example, closed world reasoning in the suppression task, or a different interpretation of falsity in the two-rule task. While discussing these possible explanations of subjects’ behaviour, the obvious criticism was put aside, namely: why different interpretations instead of the single interpretation intended by the experimenter?

The answer that will be developed in the following chapters starts from the following observation: every kind of ’given’, ranging from relations between events in the world (where causality is not given but imposed) to discourse and reasoning suffers from underdetermination. This is a broadly Kantian theme, but one supported by much work in cognitive science. In a famous article entitled ‘Going beyond the information given’ Bruner

[**5**] Brunner regards this to be the hallmark of cognition: ‘the most characteristic feature of mental life, over and beyond the fact that one apprehends the events of the world around one, is that one constantly goes beyond the information given’ [**5**, p. 218]. The article dates originally from 1957, but its main message is still intact, also in an area of considerable interest here, discourse processing (see for example Hagoort and van Berkum [**19**]).

Bruner proposes that the mind has one characteristic strategy to go beyond the information given: ‘We propose that when one goes beyond the information given, one does so by virtue of being able to place the present given in a more generic coding system and that one essentially reads off from the coding system additional information [...] [**5**, p. 224]’. In the case of a reasoning task, the coding system would have to accomplish at least the following. The connections between the premises are not given, so a common coding system for the premises is necessary to induce such connections. Likewise, the connections (if any) of the premises to other knowledge is not given; for example, what aspects of the meaning of the premises play a role. This requires a coding system in which a distinction is made between logical and non-logical constants. What coding system is applied is influenced by context, and this leads to individual differences in interpretation.

The coding system most relevant to inference is what will be called here *logical form* following Stenning and van Lambalgen [**44**, Chapter 2]. Traditionally, this phrase has been taken to refer to syntactic structures in a given formal system, usually classical first order logic. Stenning and van Lambalgen’s usage of the term is much broader however, and logical form comprises at least the following elements:

(1) Definition of a formal language
(2) Definition of a semantics for the language, that is, a definition of ‘structure’ and of a satisfaction relation, relating formulas and structures
(3) Definition of validity

In each of these three domains very different choices can be made. Under 1., there are such diverse possibilities as a fully recursive definition of the language versus the much more restrictive language of closed world reasoning. Under 2., structures may be classical first order models, but also information states. What is particularly interesting in Stenning and van Lambalgen’s definition is that validity appears as a separate entry under 3. In the classical Bolzano -Tarski definition of logical consequence, validity is in fact entirely reduced to semantics via the condition: an argument is valid if the conclusion is true, whenever the premises are true. However, many objections have been voiced against this definition (see e.g. Etchemendy [**12**], Prawitz [**34**]), which point out that truth and validity are independent though related notions. Also, in closed world reasoning validity is independent from truth, since the validity of an argument is evaluated on a special class of models only, the so-called minimal models.

‘Going beyond the information given’ plays a role in almost all cognitive processes, including those of a largely automatic character such as vision. In these cases the coding system is hard-wired, therefore, questions of normativity make little sense. One enters a completely different world when studying reasoning. It is true that experimenters administering, say, some form of the selection task, have typically assumed that only one coding system is applicable: classical logic. But alternative possibilities abound, and therefore reasoning is necessary to establish an interpretation. This form of reasoning, Stenning and van Lambalgen call ’reasoning to an interpretation’; this process involves establishing a logical form, or at least part of a logical form. In fact, the interpretation is the logical form specialised to the situation at hand.

Logical form makes it possible to connect the elements in the 'given' in such a way that further processing (i.e. inference here) can take place. Without such a common representation, inferences from a set of premises would not be possible.

CHAPTER 3

Rationality and normativity

1. Theoretical discussion: two conceptions of rationality

Referring to modern philosophical and psychological literature, two prominent conceptions of rationality can be distinguished. According to the first, rationality regards the relation between one's system of beliefs and one's actions. For instance, in the MIT Encyclopedia of Cognitive Science [**55**] 'rational agency'[1] is defined as a coherence requirement: 'the agent must have a means-end competence to fit its actions or decisions, according to its beliefs or knowledge representations, to its desires or goal-structure'. Thus, according to this definition, certain psychiatric patients (e.g. schizophrenics) would not be characterized as rational agents.

The term 'fit' here has a logical component. There is no 'fit' when an action is performed that is not part of a derived plan to achieve a certain goal. This means that taking my bicycle keys with me when I plan to take the tram is not rational, neither is carrying my driving license with me. Translating this into the context of reasoning tasks would mean coherence between the logic one uses (the norms one applies) in solving a reasoning problem and the solutions one thereby provides. Suppose a well-known mathematician takes it that he can prove a theorem only if he does not eat for one whole week. Conditional on his beliefs, the plan to starve for seven days is reasonable, and thus the agent should be seen as exhibiting a (relatively) rational behavior when he does so.

The second conception of rationality regards the conformity to an already given set of norms. This is the conception of rationality one finds in Wason [**53, 54**], or for that matter Piaget [**31**], who was the target of Wason's selection task. Whereas the first conception invites the problem of relativism, the problem with the second conception is clearly the justification of norms. In restricted contexts, like in reasoning experiments in which a particular logical form is trained and tried out (say, instruction of logic in an academic environment) this problem of justification can be momentarily suspended. However, when one is confronted with generalised conclusions about the general reasoning standards of either one person or the general population, the question cannot be avoided. This is what happens, for example, with Wason's results.[2]

To distinguish the two conceptions of rationality, one can refer to normativity as *internal* or *external* to the system, depending on whether coherence is promoted, or rather conformity with fixed norms. Whether fixed or context-dependent, in the absence of any normative constraints the concept of 'rational agency' becomes vacuous. Of course, the request for conformity to a given set of norms does not override the demands for internal coherence. An intelligible system of norms will trivially satisfy consistency and coherence.

[1]The entry is written by Christopher Cherniak.

[2]The problem of justification of norms should not be conflated with the much disputed philosophical question of whether the norms are primarily socially motivated or not. It is partly the aim of this work to argue that one needs some account of normativity in order to explain reasoning, and that an intelligible grasp of normativity can be achieved only if the reasoning norms are conceived of as *a priori.*

It is interesting to note that ”rationality” is typically used in psychology of reasoning in order to characterise a certain competence of the subject, therefore allowing for performance errors. In other words, irrational behavior is considered when there is a systematic deviation from the norms. As an example, let us consider Wason’s selection task. Wason attributed the A,4 answer to irrationality: a systematic deviation from the (here externally defined) norm, and not a momentary lapse [**53, 54**]. The deviation was taken to be systematic because the percentage in the subject population of this choice of cards was around 45%. Following this line of thinking, one would talk of a performance error if the percentage of correct answers (A,7) would be in the order of, say, 90%. However, rather than evidencing irrationality, a systematic deviation from a norm can actually indicate that the standards of rationality are not properly defined. Furthermore, what may at first seem a performance error may later be attributable to individual differences in interpretation.

2. Why interpretation matters: cognition and cognitive tasks

Matters of interpretation are central in assessing the rationality of subjects who take part in reasoning tasks. For one thing, the assessment of rationality is bound up with the issue of whether a given discourse has a unique meaning that is to be discovered. As an example, consider once more Wason’s selection task and, in particular, the experimenter’s instruction not to choose unnecessary cards. Motivated by this instruction, subjects who perceive the 7 card as potentially falsifying, might refuse to choose the 7 card because they judge the outcome 7/K to be uninformative[3]: one still needs to turn the A card. In other words, a card choice is deemed necessary if and only if every result of turning the card is immediately relevant. In this line of thinking, 7/K is presumably not immediately relevant, because one still needs A/4. Non-selecting the 7 card can then be seen as a consequence of the failure to grasp the intended meaning of the experimenter’s instructions. This is hardly irrational: the subject strives to be internally coherent, and only fails to be externally coherent because of a failure in communication with the experimenter.

When deviation from the expected answer is fully explainable, it is problematic to call this deviation a mistake. A good example to illustrate this is a Wason-type experiment reported in Stenning and van Lambalgen [**42, 44**] with the rule replaced by ‘if there is a vowel on the *invisible back*, there is an even number on the *visible face*’ (emphasis added). Stenning and van Lambalgen have observed that quite often the implication is reversed here, possibly due to ‘pragmatic normalization’. This is a procedure that is in principle quite useful to repair errors in natural language, but which generates mistakes when literal meaning is concerned. It follows that a quick identification of performance errors with errors that are usually not made is mistaken (the emphasis is on ‘usually’). In this experiment, the task material is such that it invites this so-called performance error much too often: just as in the visual domain particular two-dimensional material invites a three-dimensional interpretation because of how our visual system is constructed. Could one say that this reversal is plainly wrong, even though it is fully explainable why people make it? In other words, could one claim such an interpretation to be irrational? Even though not intended by the experimenter, the reversed or even bi-directional interpretation of the conditional is possible in some contexts, and in order for the subject to give the right answer, she might first have to ascertain that the experimenter has not made a mistake. Put differently, the concept of ‘mistake’ belongs to ordinary discourse and not to an experiment. Thus, the necessity to incorporate (or invent!) a context in order to process sentences makes the concept of rationality difficult to apply.

[3]The expression ‘7/K’ means: 7 on the visible face, K what one sees after turning the card.

The focus here has been on reasons that can justify the subject's failure to provide the right answer. However, this should not be taken to imply that interpretation issues can be safely ignored when one is faced with 'correct answers'. Indeed, there is a certain asymmetry between rationality and irrationality, in the sense that, superficially speaking, rationality is considerably easier to ascribe than its counterpart. When, for instance, the behavior of the subject is measured against well-formulated standards, e.g. a set of answers as the solution of a reasoning problem, it is only natural to take 'correct' answers as immediate proof of 'rational behavior'. However, it can always be the case that the 'correct' answer has been given for the wrong reasons. An amusing example of this occurred in Stenning and van Lambalgen's experiment on the selection task, in which a subject gave the correct answer A and 7, but, as it turned out, because she thought 7 is an even number [**44**, Section 3.6]! As this example shows, however, it is in principle feasible to detect that the 'correct' answer is not supported by the intended reasoning process by means of extensive tutorial dialogues. By contrast, ascribing irrationality in the case of 'wrong' answers is a considerably more difficult task. For the claim of irrationality to be fully justified, the experimenter would have to not only prove that there is lack of conformity with internally or externally defined norms, but also that there is no alternative justification available for the reasoning behavior of the subject. The latter is obviously an immensely more difficult task, if realisable at all.

3. Logic into the picture

A logical component is inherent in the notion of rationality however the latter is conceived. If rationality is seen as an internal to some system type of property, then the burden lies on notions like coherence and consistency between one's mental states, such as beliefs, knowledge representations, intentions, and one's reasoning behavior. If rationality is externally defined with respect to a given set of rules, the logical component is to be found in measuring one's reasoning against well-formed standards. As observed earlier in this chapter, the two conceptions distinguished here are by no means mutually exclusive.

Seeing rationality as a general phenomenon, i.e. a phenomenon that takes place over time, rather than one describing individual atomic cases, means that to formalise it one needs to account for non-monotonicity. To illustrate this, let us extend the scope of rationality to a general theory of human agency, as is done in the MIT Encyclopedia of Cognitive Science [**55**] entry on 'rational agency'. If an action is performed which is not a necessary part of a plan derived to achieve a given goal, then there is no fit between the agent's actions and plans, and if this misfit systematically obtains, then the person is judged to be irrational. Suppose now that the logic adopted here is classical, and that some theory of causality like the event calculus is invoked to construct plans. Classical logic gives a multitude of plans corresponding to the infinitely many different models of the event calculus [**52**] in which the goal is achieved. Each plan by definition fits the goal, but of course most of these models contain events (and statements about their causal effects) about which it is hard to have justified beliefs, because these events lie in the future and are not connected by causal laws to events known to have occurred. It is only natural that over time knowledge is acquired which has as a consequence either an update of the set of justified beliefs or even a revision of beliefs previously thought to be justified. These considerations have mostly negative import in showing that classical logic cannot be all there is to modelling human rationality, without by themselves proving that there is an alternative non-monotonic system, or indeed any logic at all, that provides norms (in particular rules) to go by. Positive arguments for the necessary role of logic will be supplied in the coming chapters.

CHAPTER 4

Constitutive and regulative norms

Verbal reasoning tasks are presented as pieces of discourse that hopefully do not need more than the subject's reasoning process to generate an answer. As it stands, there is nothing wrong with this expectation. However, it is an unstated assumption of almost all work on psychology of reasoning that verbal reasoning tasks can be designed in such a way that the experiment tests only what the experimenter designed it to test, and not for example natural language comprehension. The hope is that semantic interpretation of the task materials is completely determined by those materials. What is problematic with this unstated assumption is that it completely misunderstands the nature of language comprehension. However predictable the interpretation is, strictly speaking, the meanings intended by the experimenter are not furnished by the experiment itself. The reader has no other choice but to reconstruct the writer's intended meaning using her own semantic and world knowledge. We will see that clearing up this misunderstanding has important consequences for the conception of normativity.[1]

It is a common observation that there is no one-to-one mapping from words to concepts. However, applying this observation to some critical words in standard reasoning tasks, like Wason's selection task, quickly yields dramatic consequences. For ease of exposition, the task (first given in Chapter 2) is repeated here as figure 1.

Note that this is all the information that the subject is provided with. No further explanations are provided for pivotal concepts like the conditional 'if.....then', the pair 'true, false', and, more subtly, what it means 'to turn only the card you have to turn'. This will be spelled out in some detail starting with the conditional.

Wason intended the subject to conceive of the conditional as the material implication, but there are numerous other interpretations possible. An easy example of such a interpretation is the bi-conditional based on the material implication, and this possibility has been

[1]The following discussion is adapted from Stenning and van Lambalgen [**44**, Chapter 3].

> Below is depicted a set of four cards, of which you can see only the exposed face but not the hidden back. On each card, there is a number on one of its sides and a letter on the other.
>
> Also below there is a rule which applies only to the four cards. Your task is to decide which if any of these four cards you *must* turn in order to decide if the rule is true. Don't turn unnecessary cards. Tick the cards you want to turn.
>
> **Rule:** *If there is a vowel on one side, then there is an even number on the other side.*
>
> **Cards:**
>
> | A | K | 4 | 7 |
> |---|---|---|---|

FIGURE 1. Wason's selection task

considered in the literature [**16**]. Still, natural language furnishes many more interpretations. The conditional may allow exceptions, in which case the task is impossible to solve. For example, is a particular card showing A and 7 an exception or a counterexample? Instances of this interpretation as allowing exceptions can be law-like or generic interpretations of the conditional, which have the additional feature that their domain is generally indefinitely large, and not the finite set of the given four cards. Without any context the subject will have difficulty in deciding which of these interpretations of the conditional is meant. The discourse itself may provide some clues; for instance, the fact that the four cards are given may suggest that, after all, the rule doesn't refer to an indefinite population of cards. Indeed, subjects display reasoning towards this interpretation when they ask the experimenter whether the four cards depicted are really all the cards that there are. But this only amplifies the point made in Chapter 2 that coming to some interpretation requires a reasoning process.

Thus far, only one expression has been considered, namely the conditional 'if, then'. Matters quickly get worse if other expressions are taken into account as well, for instance, the pair 'true, false'. Wason intended these to stand for the classical truth values, which satisfy bivalence, i.e. the property that non-false equals true. Thus, if the conditional has no counterexamples, it must be true. However, many other interpretations are possible here as well. In connection with the law-like reading of the conditional mentioned before one may note that not having counter-examples does not yet yield the necessity that the true law-like conditional brings with it. Again, this difficulty has been observed in subjects. For instance in the two-rule task discussed in Chapter 2, some subjects[2] judged the import of the card 3 in line with the classical interpretation, i.e. they recognized that turning the 3 card would show which rule is false. However, they refused to conclude from this that the other rule therefore must be true, even though this is given as a background condition. In cases such as these the truth-values true and false are therefore independent, against the intentions of the experimenter. It may also happen that false takes on a special meaning when applied to the conditional. As mentioned in Chapter 2, Fillenbaum [**13**] observed that many people interpret a statement like '$p \rightarrow q$ is false' as '$p \rightarrow \neg q$'. In the context of the selection task, this means that a subject having this interpretation has to choose between the rules '$p \rightarrow q$' and '$p \rightarrow \neg q$'. Clearly in this case the choice of the p card (i.e. 'A') suffices to make the decision. An even more remarkable interpretation of true is as holding in almost all of the cases. In other words, for these people, true entails high probability, and this again encourages an interpretation of the task as referring to a large population of cards. The verbal formulation of the task once more provides preciously few indications of which interpretation of true and false is intended.

These two examples concern expressions that occur overtly in the formulation of the task. But contrary to first impressions it is not any clearer what the subject is asked to do once he has come to an interpretation of the conditional and its relation to the cards shown. True, the subject is given an instruction of sorts: check those cards that you have to turn in order to determine whether the rule is true or false. The difficulty with this instruction is that it is entirely hypothetical since no cards can actually be turned. This leads many subjects to think that they should come up with what may be termed a reactive plan, i.e. a choice of cards where a particular choice may be dependent on the outcome of a previous choice. Concretely, the decision to choose the 7 card may be dependent on what is on the other side of the A card. Nothing in the instructions excludes this possibility. The experimenter intended something else, namely that the decision to choose a card is not contingent upon the outcome of turning another card. However, for most people this conflicts with the other instruction not to turn unnecessary cards.

[2]In Stenning and van Lambalgen's experiments.

A different set of issues arises when we consider the suppression task. In all reasoning tasks where premises are given and the subject is asked whether or not a particular conclusion follows "necessarily", it is of paramount importance to somehow integrate the premises. But the premises given in the suppression task present considerable difficulties for this integration. Consider the standard example 'if she has an essay, she studies late in the library', the categorical premise 'she has an essay', then the additional 'if the library is open, she studies late in the library'. When the additional conditional is presented, the task form requires that it is integrated with the previous two premises. However, for all subjects except those with considerable logical training, integration here means judging the impact of the additional condition on the first two premises. Representing the task in a classical logical form, i.e. applying MP monotonically, would lead to the legitimate question of why the additional conditional was supplied in the first place. Closed world reasoning with representation of exceptions allows the expression of relevance of the additional conditional to the main conditional. Thus, the reasoning task is so constituted that it allows the integration, whereas classical logic does not allow this direct integration. The subject rightly views the second conditional as supplying relevant information and looks for a logical form that can take this into account. Thus, one main difference between the selection task and the suppression task is that whereas the former contains a number of problematic expressions, in the latter the mode of integration is left underdetermined by the task instructions.

1. Norms as constitutive of interpretation

The preceding considerations can be put in the following stark form: in an important sense the subject is confronted with a task that does not exist, and for her the most important component of the reasoning task might well be to create the task for herself in the first place. As the above examples show, this may involve getting clearer about the meanings of key expressions involved, and, at a meta-level, understanding what the experimenter wants her to do. In the case of the selection task the course of action asked of the subject is highly undetermined by the verbal instruction, and in the suppression task the inference process cannot really get off the ground before the subject has decided how the premises must be integrated.

This 'creating the task' will be refered to here as a constitutive process, borrowing this term from Kant who devoted much of the *Critique of Pure Reason* [**27**] (henceforth abbreviated to *CPR*) to showing that assuming such a process is necessary to explain the possibility of cognition. A quote from *CPR* outlining Kant's general program is a good starting point:

> l call that in the appearance which corresponds to sensation its matter, but that which allows the manifold of appearance to be intuited as ordered in certain relations I call the form of appearance. Since that within which the sensations can alone be ordered and placed in a certain form cannot itself be in turn sensation, the matter of all appearance is given to us a posteriori, but its form must all lie ready for it in the mind a priori, and can therefore be considered separately from all sensation [**27**, A20/B34].

Before comments on this passage are provided, note the intended analogy: matter corresponds to the task materials, and form to the logical form that the subject imposes on the task materials. As Chapter 2 argues, logical form is not, or is at most partially, to be found in the materials, and it is, therefore, *a priori*.

The passage just quoted implicitly assumes that there is something 'which allows the manifold of appearance to be intuited as ordered in certain relations'. A famous example will help to clarify this assumption. Hume argued that perception does not give us causal relationships, but only events succeeding each other. Kant agreed, but added (i) that awareness of causal relationships is contributed by cognition, not by perception, and (ii) that cognition must do so in order to create an objective world: for in the objective world, temporal sequences are irreversible, unlike sequences of actions generated by the subject, which are generally reversible. Causality is one of the forms of cognition, which might actually be viewed as a particular logical form. The event calculus of van Lambalgen and Hamm [**52**] proposes a formalisation for such a logical form. Causality is a constitutive principle, meaning a necessary condition for the possibility of experience. The process by which events are combined in a causal relationship is called synthesis, a notion about which more will be said in the next chapter. The focus here is placed on the consequences that the view of cognition just outlined imply for normativity.

Constitutive principles are viewed here as norms, to be called constitutive norms. Applied to causality, the constitutive norm says: ‘if you are to have experience at all, events must obey the principle of causality’. Below, the passage in CPR that comes the closest to providing a definition of what constitutive principles are:

> the dynamical laws we are thinking of are still constitutive in regard to **experience**, since they make possible *a priori* the **concepts** without which there is no experience [**27**, A664/B692].

Constitutive norms are not about regulating behaviour (which will be discussed shortly), but rather about existence: they create a phenomenon or behaviour. That a constitutive process has a normative character of its own is also observed by Searle [**40**]. Here is an interesting passage:

> Suppose, also, that in my athletic circle football is a game played according to such and such rules. Now, the specification, ‘They played football’, cannot be given if there were no such rules. It is possible that twenty-two men might go through the same physical movements as are gone through by two teams at a football game, but if there were no rules of football, that is, no antecedently existing game of football, there is no sense in which their behaviour could be described as playing football [**40**, p.35-36].

Consider a game of cards. Any specific game has to be played according to some rules, otherwise distinguishing between different games would not even be possible. These rules define the game in the sense that a play not abiding by these rules entails not playing the game, or playing a different one. The constitutive rules, as it were, create the game.[3]

Searle’s usage captures part, but not all, of the meaning of a constitutive norm that is intended here. For Searle, a subset of the rule-like felicity conditions for the performance of a speech act is required for the act to be identified as such in the first place. Quite similarly, a verbal reasoning task needs many rules to bring the task into existence for a subject. However, while the concept of a game is not constrained by a single definition of what a game should be (recall also Wittgenstein’s paradigmatic notion of family resemblance as the concept of game), inference tasks are much more constrained. In fact, it has been argued here that inference in the sense of ‘going beyond the information given’ always requires a logical form. This imposes a constitutive norm on a higher level, namely that the rules which bring a reasoning task into existence must be of the form of fixing the

---

[3]This observation has been particularly influential in the philosophy of language. According to Searle: ‘the semantic structure of a language may be regarded as a conventional realization of a series of sets of underlying constitutive rules, and that speech acts are acts characteristically performed by uttering expressions in accordance with these sets of constitutive rules [**40**, p.37]’.

parameters in a logical form, for example the notion of truth. This brings the present use of the term ‘constitutive’ much closer to a Kantian usage.

Returning to reasoning, questions about the rationality of the subject do not arise before it has been ascertained how the task is constituted by her. However, once a task has been constituted, some inferences are valid and others are not. This yields additional norms of a different nature, termed here regulative norms.

2. From constitution to behaviour: regulative norms

There is a distinction in ethics between ‘justifying a practice and justifying a particular action under it’ to which Rawls [**36**, p. 3] first drew attention. Rawls defines ‘practice’ as ‘a sort of technical term meaning any form of activity specified by a system of rules which defines offices, roles, moves, penalties, defenses, and so on, and *which give the activity its structure*. As examples one may think of games and rituals, trials and parliaments.’ (ibid, my emphasis).

Rawls argues that the legislator is in the business of justifying practices, whereas it is the judge’s task to justify particular actions, such as punishment. Rawls’ remarks point out that there is a derivational relationship between these two forms of justification (and note that typically the justification of a punishment cannot be completely derived from the practice; and also that justifying a practice is viewed as constitutive (see emphasised phrase). Similarly, ‘justifying a particular action’ is an example of the application of a regulative norm, which tells one what to do and what not to do.

Searle formulates the distinction between constitutive and regulative norms as follows:

> Regulative rules regulate a pre-existing activity, an activity whose existence is logically independent of the rules. Constitutive rules constitute (and also regulate) an activity the existence of which is logically dependent on the rules [**40**, p.34].

The question now arises: what does the distinction mean for reasoning? Constitutive norms do not capture all there is, because they establish at best only a logical form, including a general definition of validity; they say nothing about which particular inferences are valid or invalid; this has to be derived. It is regulative norms that are responsible for the distinction between valid arguments and fallacies. The reason that this distinction cannot be made on the level of the constitutive norms is that it makes sense to talk about fallacies only from the perspective of a fully specified interpretation. In other words, the main idea is that regulative norms are properly speaking always conditional on constitutive norms. However, the precise form of this conditional is not easy to state. This issue will be taken up in later chapters. One’s first idea might be to formalize the conditional as follows:

Logical form $\Longrightarrow$ inference X sanctioned by the logical form

This indeed works for monotonic logics, but it fails for closed world reasoning; there, what inferences are allowed depends on world knowledge, and moreover knowledge of your knowledge, as seen in Chapter 2. In this case the form of the conditional is rather a meta-statement such as:

closed world reasoning $\Longrightarrow$ ‘infer what follows from the completion of your knowledge’.

Any particular instance of the consequent of this regulative norm can be sanctioned given assumptions about the reasoner's knowledge; and in this sense the regulative norm itself is justified.

CHAPTER 5

Synthesis

The previous chapter concluded that when presented with the experimental material, the subjects have no choice but to reason to an interpretation. Also, that they do so by integrating the data under one logical form. The process that the subjects go through is very much reminiscent of the Kantian notion of 'synthesis'. Reasoning to an interpretation is a synthetic activity which yields information that a) goes beyond what is actually given, and b) gives rise to norms that allow for inferences to be drawn. To clarify the meaning of synthesis, it is only natural to turn once again to the *CPR*.

> By synthesis in the most general sense... I understand the action of putting different representations together with each other and comprehending their manifoldness in one cognition. Such a synthesis is pure if the manifold is not given empirically but a priori (as is that in space and time) [**27**, A77/B103].

> . . . all combination, whether we are conscious of it or not, whether it is a combination of the manifold of intuition or of several concepts, and in the first case either of sensible or non-sensible intuition, is an action of the understanding, which we would designate with the general title synthesis in order at the same time to draw attention to the fact that we can represent nothing as combined in the object without having previously combined it ourselves, and that among all representations combination is the only one that is not given through objects but can be executed only by the subject itself, since it is an act of its self-activity [**27**, B130].

What is stressed in this quote is that synthesis has a prominent *a priori* character.[1] This is in line with the previous observation that the integration of the data under a logical form is something imposed by the subject on the experimental material and not what that the material yields in and by itself. By describing synthesis as a mode of combination, the quote also motivates certain questions. In particular, is synthesis a compositional activity? Also, is 'integration' the only key notion to characterise this activity, or does some sense of 'selection' need to be assumed? These questions will be posed in the first half of this chapter.

Kant took synthetic judgements to be the end product of the mental activity of synthesis in order to contrast them to analytic judgements which correspond to an analytic process. However, it is useful to disentangle the two, namely (cognitive) process and (verbal) product, mainly in view of some enduring open questions regarding how they precisely relate.

[1]The very interesting book by Béatrice Longuenesse, *Kant and the capacity to judge* [**28**] makes the relation between synthesis and analysis hinted at in this passage the cornerstone of its interpretation:

> Indeed, one could summarize the argument [of the first *Critique*] as follows: consider the forms of the *analysis* of what is given in sensibility (the forms of "comparison, abstraction, reflection" – the logical forms of judgement) and you will have the key to the forms of the *synthesis that must occur prior to analysis*, namely the synthesis required for the sensible representation of the x's that can be reflected under concepts according to the logical forms of our judgements.

For example, is it possible to draw an absolute distinction between synthetic and analytic products of our cognition by looking only at the products themselves? Furthermore, are synthesis and analysis complementary notions, and if so, in what sense? These questions will be the focus of the second half of this chapter.

1. Synthesis and logical form

The notion of synthesis is used here to characterise a cognitive activity that enables interpretation, and ultimately makes it possible for inferences to be drawn. By itself, this use of the notion is by no means new. Two examples are provided below, which will turn out to be important.

Hintikka, in a chapter of [**23**] entitled 'Kant vindicated' argued that one can make sense of Kant's assertion that geometric truths are synthetic *a priori* by redefining synthesis: a true geometric statement is synthetic if it can only be proved by means of a construction in which more objects occur than are posited in the enunciation of the theorem. This happens for instance in the proof that the sum of the angles in a triangle is equal to two right angles; here one draws an auxiliary line through the top vertex of the triangle, parallel to the base. It is only in this extended configuration that the data, the sizes of the three angles, can be used to full effect; one must as it were go beyond the data.

There are more radical ways of going beyond the data, however. Examples of this can be found in vision, where the task of the visual system is to transform the two-dimensional retinal array into three-dimensional objects. The edges of my laptop that I now perceive are not literally there in the data, which consists of intensity values only, but have to be extracted by a laborious computational process, involving a 'coding system' in the sense of Bruner [**5**]. This is an example of synthesis which is different from the geometric case, in which the components of the product of synthesis can already be found in the data.

Where the approach taken here possibly diverges from the standard literature is that our concept of synthesis departs from at least two simple intuitive ideas typically associated with it. These ideas are: a) that the constituents of the synthesised product must exist prior to the composition itself, and b) that the product is actually synthesised from its components. The reason for us to have to depart from these ideas is that the relation between constituents and composition takes on a very different meaning in the present discussion. Roughly put, the components that support an inference (e.g. a conditional) exist only from the perspective of a particular interpretation, that is, only from the perspective of the synthesised whole. At the same time, an interpretation is not synthesised from its components, but is the outcome of much more global parameter-setting, i.e. the assignment of a logical form.

There are both cognitive and logical reasons that make synthesis a non-compositional activity. On the cognitive side, one can observe that reasoning to an interpretation is not a linear process, according to which, for instance, the subjects first perceive the data, then attribute meaning, and finally choose the pertinent logical form. The data have no semantic significance before the interpretation process gets off the ground, and this point has been stressed a number of times already in the course of this work. No reasoning task, however predictable, can be taken to exist in and by itself. However, this should not be taken to imply that the subject starts from a blank mental slate every time an interpretative process is called for. In fact, this never happens in cognition. Thus, it is plausible that subjects often start assigning a logical form even before making sense of the full set of data, that is, only with the hope that further understanding of the material will eventually be accommodated to what is already there. Revising a logical form in favor of another could be caused by exactly this, namely the failure to accommodate the full set of the experimental data under the 'default' choice of a logical form.

One can observe the phenomenon just described in the Wason selection task (see beginning of Chapter 4). The second paragraph of the task starts with the sentence 'also below there is a rule which applies only to the four cards'. Some subjects assign a logical form to the task based on this sentence and take the rule to be true. Of course this goes against what is said in the sentence immediately following this one: 'your task is to decide which if any of these four cards you must turn in order to decide if the rule is true'. This sentence asks the subjects to establish whether the rule is true. Some subjects experience difficulties in incorporating this new information, and at times they verbalise their puzzlement, for example as follows: 'there is something in the syntax which I don't understand. It says here that the rule is true, isn't?'.[2] Clearly, taking the rule to be true means that the subject can only choose cards which possibly verify the rule. This makes, in particular, the 7 card irrelevant.

There are additional logical reasons that make synthesis a non-compositional process. For one thing, under-determination is a serious problem. For synthesis to be a compositional activity, one should be able to have a grasp of what the components, i.e. the atomic meanings, in this combinatorial procedure are. According to what has been established in the previous chapters, the data cannot play the role of these atoms, since they lack meaning prior to the assignment of a logical form. However, the components of logical forms themselves, that is, the setting of the parameters, cannot play this role either. This is obvious in closed world reasoning, where world-knowledge has to be computed. In non-monotonic reasoning, 'going beyond the information given' is not restricted to the way the subject understands the experimental material in a reasoning task. Rather, it also occurs throughout the process of arriving at an interpretation, since adding information might lead to conclusions not previously accepted or derivable from the hitherto existing logical form. What one ends up with in closed world reasoning may be far removed from where one started. Thus, it is fair to say that what constitutes an interpretation does not exist prior to assigning a logical form. In a paradoxical way, the components of a non-monotonic inference exist only in so far as the composition exists too.

One would think that, by contrast, having a fully specified classical logical form makes for a fully compositional process of interpretation. This would turn synthesis into an uninteresting notion. Of course, any interpretation is synthetic if only in that it is grounded on a translation of the data into a formal language, thus integrating the data. But what does synthesis involve when the logical form chosen is that of classical logic?

The answer to this question requires a more careful look at the process of synthesis. So far, synthesis has been discussed mostly as a process of combination, but it must not be forgotten that it also involves a good deal of selection. This is true for synthesis in cognitive science as well. The informational situation of human beings is slightly paradoxical: there is both too little and too much. One way cognition solves the problem of 'too much' is via categorisation: representing continuous input in the form of discrete categories, as happens for instance with colours. This is a form of selection: to a certain degree precise wavelengths of colours are unimportant. Another example is furnished by phonetics. Initially babies can process speech input in more or less continuous fashion and they can make distinctions among phonemes that are not in their language. Later on, categories become hardened, and children lose the capacity to make distinctions that do not occur in their native language.

[2]Further data corroborating this point can be found in [**41**, p. 299].

The first thing to observe is that imposing a classical logical form happens less often than one might expect. If ‘closing the world’ boils down to eliminating abnormalities considered to be irrelevant as well as alternative rules, there may be a temptation to understand classical logic as an ‘opening the world’ type of reasoning. Reasoning classically would then mean that one is always ready to consider exceptional circumstances whether these are mentioned or not. However, can we really say that one is applying classical logic when taking into consideration concrete information, whether in the form of abnormalities or of alternative rules? This issue will be discussed using AC as an example: from q and $p \to q$ conclude p.

The question to ask here is the following: if one suppresses, or, (more neutrally) does not endorse, AC in the presence of only two-premises, is it because AC is always fallacious in classical logic or because there has been found an example (or it is likely that one will be found) of a true instance of q and a false instance of p? Let us consider AC applied to ‘if the key is turned, the car starts’ and ‘the car has started’[3]. Closed world reasoning derives the conclusion ‘the key was turned’, while in classical logic this inference - based on these data only - is not allowed. The question is whether this non-inference is justified by saying ‘in classical logic, AC is not valid because $p \to q$ is true if q is true and p false’, or rather by saying ‘there are other possibilities for starting a car, like jump starting’. In the latter case, one could say that the world is opened selectively, to allow a real possibility not hitherto considered. But this means that one may still view non-endorsement of AC as an application of closed world reasoning for rules; only the set of rules turns out to be larger than previously thought. The point here is that, however large the set of rules considered, it still constitutes a closure of the world. This is because one considers not the abstract truth values of classical logic, but rules and abnormalities in the world, therefore, some selection (therefore, some form of synthesis) has been performed regarding which of the alternative rules and abnormalities are sufficiently relevant to be mentioned. Accordingly, it seems that, if to suppress AC it does not suffice to say ‘p can be false even though $p \to q$ and q are true’, and instead concrete instances have to be supplied of p and q having this property, then one has already taken leave of classical logic.

Engaging in a classical interpretation seems then to be a straightforward activity, one in which the subject does not go beyond the premises given, and synthesis manifests itself only as a form integration. But it is precisely the absence of substantial selection that accounts for the processing difficulties of classical logic. Cognitively speaking, having to suppress world-knowledge is an all but effortless process. Also, a logical form which encourages a literal interpretation tends to come across as less natural than one which exploits implicit or non-stated information. In fact, if Wason type results show anything, it is that classical thinking requires conscious, laborious effort for most of the subjects lacking the relevant academic training. This is because one has to work hard to do away with relevant issues and concentrate only on stated information as it translates into abstract truth-values. Of course, reasoning classically does not always mean suppressing relevant information; when it does not, synthesis becomes a rather straightforward compositional procedure. But this is hardly the case when one has to deal with arguments as they are formulated in natural language.

2. Synthesis: product *vs.* process

Since Kant’s main question was how synthetic a priori judgements are possible, the analytic/synthetic distinction is therefore central to the purposes of the *Critique*. According

[3]From Oaksford and Chater’s running example of the suppression task [**30**].

to Kant, the end product of the mental activity of synthesis is synthetic judgements. Following Kant, one would think that a deeper inquiry into the nature of synthetic judgements can open a window to understanding synthesis itself. However, we will see that this is not so; rather, the process of synthesis must be invoked in order to justify the distinction. This discussion will also shed light on the much disputed question of whether and in what sense logical inferences are analytic. Frege, by defining a true statement to be analytic if for its proof we need to refer only to logical laws and definitions, implicitly declared all logical laws to be analytic [**15**, para. 3]. Today one would say that logical laws are true by virtue of pre-existing meanings of the logical operators, hence again analytic. We will see that there are serious reasons to question whether this provides an informative view of logic.

Below we reproduce Kant's notorious definition of analytic and synthetic judgements which he provides in his introduction under the section entitled 'on the difference between analytic and synthetic judgements':

> In all judgements in which the relation of a subject to the predicate is thought (if I consider only affirmative judgements, since the application to negative ones is easy), this relation is possible in two different ways. Either the predicate B belongs to the subject A as something that is (covertly) contained in this concept A; or B lies entirely outside the concept A, though to be sure it stands in connection with it. In the first case I call the judgement analytic, in the second synthetic [**27**, A6-7].

This definition points to an absolute distinction between analytic and synthetic judgements determined by whether an intensional containment relation[4] holds between subject and predicate. This containment relation can be formulated in modern terms by saying that the predicate does or does not add information to the subject. Such an absolute distinction applied to the judgements themselves (and not reasoning processes) has been questioned time and again in the philosophical literature, with Quine [**35**, Chapter 2] the most famous example. In a much more detailed manner, Hintikka [**23**, Chapter VII] has argued for an analytic/synthetic distinction within the realm of logical tautologies by looking at the structure of the proofs of tautologies: roughly speaking, if the number of individuals introduced in the proof of a tautology equals the number of quantifiers[5], the tautology is analytic; otherwise it is synthetic. This is of course parallel to the discussion of the geometry example at the beginning of this chapter. What is important here is that the distinction between analytic and synthetic judgements is made on the basis of the processes producing these judgements. Thus, process has priority over product here.

The problem of relativism just identified shows that one has to look at more than just the statement itself to decide whether it adds new information or exploits given information. Also, from the perspective of cognition, which is the present interest, the distinction is particularly problematic when applied to judgements only: it is not always immediately obvious whether a predicate is contained in the subject's concept; for example, sometimes a proof is necessary. If a proof is necessary, then it becomes problematic to say that a predicate does not add information to the subject, especially in undecidable systems. This could, of course, be taken to suggest that a new definition is necessary which requires that the predicate can be immediately seen to be contained in the subject. But such a definition would in effect render the notion relative to persons, perhaps even to points in time. It would allow, for instance, a situation in which the speaker communicates an analytic judgement to the hearer which comes across as a synthetic one. Analytic judgements become then no different from known information, and the distinction does not account for a property of the statements themselves; the containment relation would ultimately depend on

[4]As opposed to the extensional inclusion relation.

[5]Assuming the formula is put in prenex normal form.

one's background knowledge, or worse, which background knowledge is salient at a given moment. Again, the distinction between analytic and synthetic judgements is reduced to a distinction between processes: whether or not there exists an immediate inference to the judgement, in a particular human being. This means that judgements can be both analytic and synthetic, depending on who is making the judgement, and in what circumstances.

Take, for instance, Kant's favourite arithmetical example '7+5=12'. The statement '7+5=12' can be viewed both as analytic and as synthetic. It is synthetic when considered as the product of the constructive ('synthetic') activity of the mathematician, which proceeds in the absence of any formal system. It is analytic if a formal system has been chosen to represent the activity of the mathematician, and the statement '7+5=12' is derived in this formal system. That is, the statement becomes analytic if all relevant parameters have been fixed before attempting to verify the statement. However, if a mathematician is presented with the numbers 7 and 5, and the instruction to add these up, and only then chooses a formal system in which to conduct the computation, the statement is again synthetic.

Ultimately, what is wrong with viewing tautologies as analytic, is the assumption that there are definite meanings of logical operators to begin with, even if one momentarily abstracts from the process leading from meanings of operators to tautologies. In natural language there are no such definite meanings. The assumption holds only for meanings fully specified in a logical form, and the analyticity of a tautology is always parasitic on that logical form. Thus logical laws can be analytic only in virtue of a preliminary synthesis. The next chapter studies in detail an approach to logic which takes this dependence of analysis on synthesis seriously.

A final question to be asked about the relation between product and process is the following: can the product of reasoning, seen as an argument form, be informative of the process that the subject goes through to get there? Put differently, is it plausible to conceive of synthesis as a product-like representation?

Some people might claim that the product of an inference is always a sentence in a fixed universal logic, which determines how to reason with the sentence. One should note here, especially to distinguish these remarks from the earlier ones regarding the ability to define argument patterns, that taking the sentence to be already part of the universal logic, its logical constants ipso facto carry meaning. In other words, the product is not to be viewed as a purely syntactic ('schematic') representation, but rather as equipped with meaning.

The relation between product and process is very pertinent to the question of normativity, and particularly to the distinction between constitutive and regulative norms that was discussed in the previous chapter. In the view just advocated, the constitutive norms play no role at all and reasoning is governed by a set of regulative norms only. This is because the inferences take place in a fixed universal logic and therefore the constitutive norms are not needed to determine what the regulative norms are.

To take a concrete example of this, consider AC: from $q, p \rightarrow q$ and the belief that there is no true conditional $r \rightarrow q$ we infer p. The data suggest that the premises are actually $q, p \leftrightarrow q$, and in a logic that claims to be universal one would expect that p easily follows. The product here is the inference $q, p \leftrightarrow q \models p$, and the process is what is implied in the expression 'the data suggest'. The later expression, one might claim, is not relevant to the inference as such. The constitutive norms supposedly governing the process of interpretation seem to be unrelated to the regulative norm licensing the inference $q, p \leftrightarrow q \models p$. So, how can we claim that a regulative norm is always determined by a constitutive norm? Is it not the case that as soon as the understanding of the logical situation is accomplished what we all naturally do is just classical logic? Alternatively, does interpretation really matter as far as the inference in itself is concerned?

Studying reasoning along the lines just sketched would be cognitively uninformative, because, although at first sight it seems so, in the end the inference in itself is not justified. The inference $q, p \leftrightarrow q \models p$ justifies drawing p as a conclusion only in view of the original data (that is, the input). This is only if the process leading from $q, p \rightarrow q$ and the belief that there is no true conditional $r \rightarrow q$ to '$q, p \leftrightarrow q$' is 'faithful'. More precisely, the data $q, p \rightarrow q$ and the belief there is no true conditional $r \rightarrow q$' are constituted as '$q, p \leftrightarrow q$ + MP as inference rule', and it is because of this relation of constitution that AC becomes a regulative norm. It follows that the product of synthesis is meaningless without an indication of what it is a synthesis of.

CHAPTER 6

Harmony

The conclusion of the previous chapter can be summed up in the following passage from the *CPR*:

> One can here easily see that this action [synthesis] must originally be unitary and equally valid for all combination, and that the dissolution (analysis) that seems to be its opposite, in fact always presupposes it; for where the understanding has not previously combined anything, neither can it dissolve anything, for only through it can something have been given to the power of representation as combined [**27**, B130].

This chapter turns to the philosophy of mathematics to find further inspiration regarding the main question: what should synthesis be like in order for rules to arise that make inferences possible? It so happens that in the debate between platonistic and constructivist conceptions of mathematics the distinction between regulative and constitutive norms has arisen, and something can be learned from the way the issue has been treated in this domain. Following a very brief introduction to constructivism, the focus will shift to an idea developed by Dummett and Prawitz that the justification of logical laws must be given in proof-theoretic terms. This will lead the discussion to the crucial concept of 'harmony' (between introduction rules and elimination rules in a natural deduction system), and of how this concept can be exploited for the purposes at hand.

1. Constructivism

The emphasis in this thesis on interpretation, together with synthesis as an integral and necessary part of it, betray a constructivist point of view: interpretation and logical form are not out there in the data to be found by the reasoner, but they arise through an active and dynamic process of construction.

Constructivism emerged in mathematics mainly as a reaction to a platonistic conception of mathematical objects, that is, a conception of mathematical objects as existing independently of the mental acts of the mathematician. This is what Brouwer characteristically calls the 'observational standpoint' [**4**, p. 1]. From such a standpoint, mathematical practice is devoted to discovering and exploring already existing mathematical facts. This view is very much like the attitude at the very opposite pole of the one adopted here, namely taking classical logic (or any 'Universal Logic') as the logic to recover the logical form in any given set of data. Taking classical logic to be the paradigmatic conceptual or cognitive framework (the universal language, the language of thought) forces one to adopt an observational point of view towards logic (cf. Frege). In contrast to this, the active process of interpretation is taken here to be part of what it means to assign a logical form. It is this point that is highlighted further here by explicitly subscribing to a constructivist point of view. Additionally, constructivism will provide some insight into the justification of logical laws.

Constructivism arose with questions of legitimacy concerning mathematical practice. Consider for instance constructivism as embodied in intuitionism, which is based on opposition to abstract, axiomatic mathematics and a simultaneous emphasis on internal, mental evidence for one's mathematical assertions. One can view intuitionism as adopting a particular logical form, namely the Brouwer-Heyting-Kolmogorov interpretation of the logical constants.[1] In this interpretation the implication, for instance, is defined as a construction which transforms a proof of the antecedent into a proof of the consequent. However, one might object, would not such a view (i.e. intuitionism as arising from a particular logical form) contradict the spirit of intuitionism, which takes logic to follow from mathematics and not the other way around? The next quote gives an example of this tendency in intuitionism:

> Logic is not the ground on which I stand ... [a] logical theorem is but a mathematical theorem of extreme generality; that is to say, logic is a part of mathematics, and can by no means serve as a foundation for it (Heyting [**20**, p. 6]).

One way to understand pronouncements such as Heyting's is via the earlier introduced distinction of reasoning to, and reasoning from an interpretation (see Chapter 2). Heyting is not arguing here that a logical form (defined as parameter-setting resulting from reasoning to an interpretation) cannot serve as a foundation; indeed, one needs a meaning for the logical operators to get started. Rather, what he objects to is the idea that logical laws, the results of reasoning from the interpretation, can ever exhaust the logical form appropriate to intuitionistic mathematical reasoning. For example, the introduction rule for the implication considers only the special case in which the transformation posited by Heyting consists in appending a piece of proof (establishing the consequent assuming the antecedent) to a proof of the antecedent. The introduction rule for the implication $A \rightarrow B$, together with the other rules, basically operate only on the information that can be extracted from the formula A, and the possibility to extract information from the proof of the antecedent is not used. Consequently, it is in principle possible that another introduction rule will be proposed which exploits the meaning of the intuitionistic implication to a greater extent. Thus, if the interpretation of the logical constants is that of Brouwer-Heyting-Kolmogorov, reasoning from that interpretation by means of the introduction rule for the implication does not exhaust the interpretation.

The conclusion of the preceding considerations is that it is not inimical to the spirit of intuitionism to view it as the adoption of a certain logical form, as long as one does not assume that this logical form can be exhausted by concrete inference patterns. In fact, intuitionistic mathematics is chosen here as an example not because of its normative claims, but because it clearly shows how logic can play a constitutive role in cognition (here mathematical cognition), and also that when logic plays this role, it often cannot be conceived of as a set of rules only.

That constructivism can be found in other forms of mathematics beside intuitionism is well-known. Posy [**32**, p.130] makes a distinction between what he calls 'constructivity of the right': formalism in Hilbert's program, and 'constructivity of the left': Brouwer's intuitionism. The existence status of mathematical objects is a key to understanding Posy's distinction. Roughly put, for constructivists of the left, only those objects exist that are constructed in a (albeit ideal) mind; a familiar example of what such a conception typically excludes is the proof by contradiction. For constructivists of the right, however, objects that exist can in principle be constructed. Notice that formulated in these terms, a platonic conception of mathematics is only incompatible with 'constructivity of the left'.

[1] Explained for instance in Dummett [**10**].

On the other hand, Hilbert was himself a Platonist and he believed that mathematics could operate just as well on the assumption that infinite entities, that is, entities that the mind cannot grasp such as the real numbers, are fictitious. Such fictitious entities he took to be indispensable in organising real entities in the sense that they can greatly simplify the proof of a theorem; but he believed that in the end one has to show that the same theorem can be proven using finite objects only (Hilbert [**22, 21**]).

Here we generalise even further, by taking a constructivist approach to mean that any imposed logical form is constructed by the subject. As a consequence, constructivism is not given here a very specific meaning, but it serves as a theoretical starting point, namely viewing interpretation as constructing an appropriate logical form. It is not specific because it does not specify criteria that allow some logical forms and rule out others, as was the case with Brouwer's or Heyting's mathematics. Instead, following the familiar Kantian theme, whatever is conceived of as a semantic representation of a reasoning problem is (by default) constructed by the mind. For example, a classical understanding of disjunction will emerge from the construction of a truth-functional representation, contrasted to an intuitionistic understanding. Similarly, the law of excluded middle will hold depending on the choice of one's notion of truth and particularly whether bivalence is endorsed.

One caveat is in order before turning to proof-theoretic justifications of inferences. In philosophy, constructivism is typically taken to imply some form of anti-realism (Dummett [**11**], and it is also ontologically loaded in mathematics. Whereas one cannot say that constructivism led Hilbert to a rejection of realism, it still holds for both 'constructivity of the left' and 'constructivity of the right' (albeit for different reasons) that objects that cannot be constructed do not exist.

Perhaps surprisingly, the problem of existence is not important here and can be safely put aside. In fact, the way constructivism features in our context is closer to a cognitive understanding, according to which meanings are entities constructed by cognition in language comprehension. Such an approach has no objections to meanings as Platonic entities. It only holds that such entities are irrelevant to cognition, and that the role of meanings in comprehension and production can only be studied when viewing them as constructed by the mind. Following this line of thought, what we are interested in are the logical properties of drawing inferences – both in general terms and with respect to specific logical forms – and not the realistic status of the objects that these logical forms entail. In particular, plausible suggestions are sought of how constructing semantic interpretations gives rise to reasoning norms.

In the course of mounting an argument for intuitionistic logic and against classical logic (and following Gentzen [**18**]), Dummett [**11**] and Prawitz [**34**] have proposed a distinction within natural deduction rules for logical constants that attempts to do the same work as the distinction proposed here between constitutive and regulative norms. The precise way this distinction is drawn by Dummett and Prawitz will prove very informative here.

2. Introduction rules in Natural Deduction as constitutive of meaning

> First I wished to construct a formalism that comes as close as possible to actual reasoning. Thus arose a 'calculus of natural deduction' (Gentzen [**18**]).

In his 'Untersuchungen über das logische Schliessen' of 1935, from which the above quote is taken, Gentzen set out to explore the syntactic properties of the logical constants for first-order predicate logic by constructing a formalism that he wished to resemble 'as close[ly] as possible actual reasoning'. Roughly speaking, the 'calculus of natural deduction' captures the way one reasons with the logical constants in terms of introduction rules, that is, rules for introducing a connective, and elimination rules, that is, rules for extracting information from a premise containing that connective as main logical sign. The result is a rule-like definition of the logical constants, according to which there is both an introduction and an elimination rule available for each of the connectives[2]. We give the rules for conjunction as an example.

(a) $\wedge$I

$$\frac{\begin{matrix}\Delta \\ \vdots \\ A\end{matrix} \qquad \begin{matrix}\Delta' \\ \vdots \\ B\end{matrix}}{A \wedge B}$$

(b) $\wedge E_{left}$ $\qquad$ $\wedge E_{right}$

$$\frac{\begin{matrix}\Delta \\ \vdots \\ A \wedge B\end{matrix}}{A} \qquad\qquad \frac{\begin{matrix}\Delta' \\ \vdots \\ A \wedge B\end{matrix}}{B}$$

Gentzen noticed that there is a special relation between the introduction and the elimination rules. One can observe the correspondence between derivations (a) and (b) above, as if they are the mirror-image of each other. At the same time there is a kind of hierarchy among the two; following the constructive (compositional) arrow (i.e. (b) placed after (a)), eliminating the connective seems in a sense to 'undo' the 'doing' of its introduction. In Gentzen's words:

> The introductions constitute, as it were, the 'definitions' of the symbols concerned, and the eliminations are, in the final analysis, only consequences of this [**18**].

This special relation between the two types of rules is further exploited by Prawitz in his work on proof-theoretic semantics. Proof-theoretic approaches typically seek to identify the meaning of logical constants with the role they play in the system of inference rules. According to Prawitz:

> . . . the rules for introduction inferences determine the meanings of the logical constant concerned, while the rules for elimination inferences are justified by these meanings [**34**].

We see here the beginnings of an analogy, where the ensemble of introduction rules may be viewed as a logical form, hence a constitutive norm. These ideas will be the focus of attention in the coming sections.

3. Proofs as constitutive of meaning

The system of natural deduction found its most articulate form in the work of Prawitz [**33**]. Starting from there, Prawitz (together with Martin-Löf) has attempted to draw out the philosophical consequences.

Prawitz relates meaning to justification, and justification takes a concrete form in the notion of proof. In the opening line of Prawitz [**34**] we read:

[2] For simplicity, and where it doesn't make any difference, I will be referring to propositional constants only.

> Since we are interested here in logical consequence, we shall focus on how one is to understand that an argument is valid in virtue of the meaning of the logical constants occurring in the sentences of the argument [**34**, p. 678-9].

According to Prawitz [**34**], Gentzen was the first to suggest that 'certain ways of proving a sentence could be seen as determining its meaning'. This is the case with example (a) in the preceding section, since what the conclusion means is what one does in order to derive it. Of course, not all actual proofs of a sentence can be seen as constitutive of its meaning. A proof like (b) with an elimination rule at the last line does not contribute to the meaning of the conclusion: what A means is not determined by the conjunction in $A \wedge B$ from which it is derived. The same holds, for instance, for $A \vee B$ in a proof whose last line is an 'all' elimination, modus ponens, or even worse, double negation elimination.

Seeing proofs as constitutive of meanings leads one to single out certain proofs as valid in view of the meaning of the main connective, namely the ones that are constructed on the basis of introduction rules. These proofs are called 'canonical' or 'direct' proofs. Speaking informally, a proof of a sentence is canonical if in its last line the main connective of the sentence is introduced. The idea is most easily explained using conjunction and disjunction as examples. One can say that a proof of $A \wedge B$ is canonical if canonical proofs are available for both A and B. Similarly, a proof of $A \vee B$ is canonical if one has a canonical proof for A, or if one has a canonical proof of B.

The canonical or direct proofs do not exhaust the constructively acceptable proofs. In fact, some non-canonical proofs are as justified as canonical proofs are. Take, for instance, the proof of A from $A \wedge B$ in (b). If one has a canonical proof of $A \wedge B$ then by definition one has a canonical proof of A. It follows that (b) can be transformed into a canonical proof, i.e. the derivation of A from itself. This gives rise to the notion of 'categorical' (also known as 'indirect') proofs. A proof is categorical if there exists an effective method that transforms it into a canonical proof.

Matters become considerably more complicated when one asks what it takes for a proof of $A \to B$ to be canonical. Clearly, the recipe used for conjunction and disjunction is of no help here: having a canonical proof (alternatively, 'evidence') of A and/or of B is not what is required in order for the implication to be introduced. Prawitz observes that one needs to rely on a different type of proof, namely a hypothetical proof, in order to specify the introduction of implication: if the implication to be introduced is $A \to B$, one says that it is assuming A that B follows. By assuming A in a hypothetical proof one assumes that in principle there exists a canonical proof of A. However, possession of such proof is not required as was the case with conjunction and disjunction.

Depending on the sentence at hand, one is very likely to apply elimination rules in order to introduce implication. This is the case when the consequent is derived from analysing the meaning of the antecedent as it happens with a sentence of the form $A \wedge B \to B$. Of course, this need not always be the derivational process to go through. One may simply append a proof of B under the assumption of $A \wedge B$ to be discharged later by the introduction of the implication. A special case of this is when a tautology is given for B, and thus the derivation of the consequent in no sense relies on the meaning of the antecedent. Unlike conjunction and disjunction, the introduction rule for implication may reflect very different roles for the meanings of the components in the inference process.

Following these considerations one comes to realise that justification on the basis of the meaning of a sentence extends beyond proofs with introduction rules only. Based on Prawitz [**34**], we can give the (simultaneous) recursive definition of canonical and categorical proofs for all propositional connectives:

- every canonical proof is categorical
- a proof of $A \wedge B$ is canonical if its last line is conjunction introduction and one has a categorical proof for both A and B
- a proof of $A \vee B$ is canonical if its last line is disjunction introduction and one has a categorical proof of A or a categorical proof of B
- a proof of $A \to B$ is canonical if its last line is implication introduction and it contains a hypothetical proof of B from A, that is, an effective method for transforming a categorical proof of A into a categorical proof of B

An example will help to illustrate these notions. Consider the tautology '$(A \wedge B) \to A$'. Its proof, conjunction elimination followed by implication introduction, has a sub-proof which is not canonical, namely conjunction elimination. But if we have a categorical proof of '$A \wedge B$', it can be effectively transformed into a canonical proof, i.e. a proof consisting of categorical proofs of A and of B, followed by conjunction introduction. The effective method postulated by the notion of hypothetical proof can then be given as follows: transform the categorical proof of '$A \wedge B$' into a canonical proof, isolate the canonical proof of A in this proof, and give the canonical proof of A as output. The proof of $(A \wedge B) \to A$ in natural deduction is a guide toward constructing the required effective transformation, but is not identical to it.

Note that by the definition of categorical proof, a canonical proof of $A \to B$ transforms canonical proofs of A into canonical proofs of B. The same idea can be used to explain what it means that there is a valid argument from premises $A_1, \ldots, A_n$ to conclusion B: there must be an effective operation which transforms categorical proofs of $A_1, \ldots, A_n$ into a categorical proof of B.

4. Intuitionism and the idea of 'harmony'

Non-canonical proofs have to be justified on the basis of the meaning of the connectives, which is in turn given by canonical proofs. The idea of 'harmony' is concerned with the form that such a justification is allowed to take.

Harmony is a notion introduced by Michael Dummett in order to characterise what it means to justify logical laws. Justifying logical laws can proceed only by giving an account of the meaning of logical operators, and Dummett opts for a verificationist theory of meaning. Given a connective O, there will be rules for asserting a sentence whose main connective is O, and rules which have as main premise a sentence with O as the main connective. Dummett postulated that there must be 'harmony' between these two uses of O:

> What is for the introduction rules and the elimination rules governing a logical constant to be in harmony? [H]armony, in the general sense, obtains between the verification conditions or application conditions of a given expression and the consequences of applying it when we cannot, by appealing to its conventionally accepted application conditions and then involving the conventional consequences of applying it, establish as true some statement which we should have no other means of establishing ... [F]or an arbitrary logical constant c it should not be possible, by first applying one of the introduction rules and then immediately drawing a consequence from the conclusion of that introduction rule by means of an elimination rule of which it is the major premise, to derive from the premisses of the introduction rule a consequence that we could not otherwise have drawn [**11**, p. 247–8].

Considerations of harmony actually arise as soon as there are several distinct ways to assert a given sentence, for instance direct verification and indirect inference. Suppose one draws an inference from true premises using an elimination rule. The conclusion is then true by assumption, but if the conclusion is sufficiently simple, it may be subject to direct verification as well. In other words, indirect inference and direct verification must yield the same results. This is the normative character of harmony. This normative character has some implications for the relation between truth and validity.

Consider first the classical notion of validity, traditionally jointly ascribed to Bolzano and Tarski: 'an argument is valid if, whenever the premises are true, so is the conclusion'. Disregarding the precise meaning of 'whenever', one sees here that validity is reduced to truth, and that hence problems of harmony cannot arise: if it is discovered by direct verification that the conclusion of an apparently valid argument with true premises is false, then the argument is simply invalid after all. But the fact that validity has not been characterised independently of truth is considered by some (e.g. Etchemendy [**12**], Prawitz [**34**]) to be indicative of its utter uselessness in inference, so this trivial way to establish harmony is not convincing. On any approach in which truth and validity are independent (although related), harmony must play a role.

In Dummett's view, truth must be replaced by provability, and validity is also characterised in terms of proof: an argument is valid if any proof of its premises can be transformed into a proof of its conclusion. In effect this means that proof theoretic validity becomes the fundamental concept in the justification of logical laws, and truth does not play any role. In an ideal logical system, every conclusion derived from an elimination rule can in principle also be derived by a proof whose last step is an introduction rule. The introduction rule corresponds to the direct verification mentioned above, so here direct verification is system-internal (as is the case with analytic judgements, see Chapter 5).

We saw how identifying the meaning of a connective with its introduction rule leads to singling out certain proofs called canonical proofs as constitutive of the meaning of sentences. Given A and B, for instance, the rule-like meaning of $\wedge$ determines the meaning of $A \wedge B$. We also saw that some non-canonical proofs, that is, proofs containing elimination rules, are as justified as canonical proofs are. The notion of a categorical proof was invoked to characterise exactly those proofs that can be transformed into canonical ones. It is precisely the canonical and categorical proofs that can be used to give substance to the notion of harmony. For the requirement that for a proposed elimination rule, a categorical proof of the premises must be transformable into a categorical proof of the conclusion, shows what Dummett demands: every conclusion of the elimination rule can be obtained directly from a proof of the premises of the corresponding introduction rule.

Conjunction is the simplest example. Asserting a sentence of the form $A \wedge B$ means that one possesses a categorical proof for A and a categorical proof of B. In other words, one has a 'more direct', indeed canonical, proof of A, without the detour of conjunction elimination. A similar argument holds for disjunction and implication. It follows that the elimination rules do not lead us outside the realm of canonical proofs. This is an important observation to make, since an analysis of sentences is often necessary as was the case with $(A \wedge B) \rightarrow A$.

Dummett claims that theories which characterise the meaning of logical constants by means of introduction and elimination rules governed by harmony, enjoy the property of compositionality, since the meaning of $A \bullet B$ (for $\bullet$ a connective) is explained in terms of proofs of A and of B, together with a specific way to combine these proofs. Dummett applied his theory of meaning – verificationist, compositional, governed by harmony – to existing logical systems, and found them wanting; in particular classical logic.

Indeed, not all proofs in natural deduction (the qualification is important; the situation is different in sequent calculi) are categorical proofs when one works classically. The idea of direct proofs determining the meaning of sentences finds its optimal expression in a natural deduction system that excludes the rule of double negation elimination. Once one allows double negation elimination, non-canonical proofs are possible that cannot be transformed into canonical ones. The most drastic example of this non-categoricity is the proof of the law of excluded middle. It suffices here to notice the following: a tautology (namely excluded middle) can be proven by means of double negation elimination for which an effective method that yields a direct proof is not available. Indeed, in the introduction rule for disjunction, if there is no proof for either A or B, one cannot introduce disjunction at the last line of the proof. A clear example of this is the atomic case $p \vee \neg p$.

There are, of course, also conceptual grounds for believing double negation elimination to be suspect: if the meaning of negation is determined by its introduction rule (via the introduction rule for the implication), then double negation elimination had better be derivable. But it isn't, as can be established via intuitionistic Kripke models. So there is prima facie little support for double negation on the present conception of meaning. This conclusion is too quick, however: perhaps the introduction rule for negation is simply too weak and must be replaced by a stronger one. This point will be amplified in a more general context in Chapter 7.

Setting these doubts aside for the moment, excluding double negation elimination from the standard system of natural deduction leads us to intuitionism, and intuitionistic logic can be shown to be harmonious. Prawitz has shown in his normalisation theorem (Prawitz [**33**]) that any proof of a tautology in natural deduction without double negation elimination is categorical, i.e. it can be transformed effectively into a canonical proof. Such a proof has the property that there are no formulas which are first introduced and immediately afterwards eliminated. If proofs with premises are also considered, one obtains the following 'sub-formula property': all formulas occurring in the derivation of A from Γ are sub-formulas of A or Γ . Furthermore the resulting proof structure looks (very) roughly as follows (the so-called 'normal form'): first the premises are reduced ('analysed') using elimination rules, then the conclusion is built up ('synthesised') using introduction rules[3]. This way, the information that there exists a canonical proof of the premises is used in mirror image to extract the relevant information from the premises via elimination.

5. Harmony breaks down

Harmony as conceived by Dummett and Prawitz has the great virtue that proofs can directly exploit the meanings of the logical constants, since these meanings are given by an especially simple kind of proof, namely the introduction rules. To explain why this is so appealing here, let us draw an analogy with concepts used earlier in the thesis. We can view the introduction rules as constitutive norms, specifying the logical form, and elimination rules as regulative norms. Here, of course, introduction rules do double duty as regulative norms as well. If the system is harmonious, the regulative norms are fully determined by the constitutive norms. This way, the Dummett-Prawitz conception of harmony seems to provide a solution to the present question, namely how to derive norms from logical form. To put this in other, Kantian, terms: the ensemble of introduction rules can be viewed as a synthesis, the elimination rules as analysis. If this analogy is correct as a general model of reasoning, then one need only ask what should synthesis must look like for analysis to be possible. Analysis in this case is the use of elimination rules that are justified by the synthetic introduction rules.

Indeed, one can see Kant struggling with the justification of the 'hypothetical inference' (i.e. elimination of implication) when he says that, given A and $A \rightarrow B$, how can it

[3]Very roughly, because valid only for the fragment of intuitionistic logic not containing the existential quantifier. See Troelstra and Schwichtenberg [**49**, p. 143ff].

necessarily follow that B? Must it not be the case that the 'ground' of this necessity lies in the necessary connection between A and B? To Kant, this seems problematic; in fact he formulates his generalisation of 'Hume's problem' (on causality) as '[how could anything] be so constituted that if that thing be posited, something else must also necessarily be posited ...' (Preface to *Prolegomena*, [**26**, p. 257] Longuenesse [**28**, p. 355] gives an interesting quote from the *Metaphysik Volckmann* (a set of lecture notes contemporaneous with the *Prolegomena*) in which Kant muses on MP and notes that while it does not present a difficulty if the connection between antecedent and consequent is analytic, it is highly problematic if the connection is synthetic.

> In the concept of a real ground there is a synthetic connection; in the concept of a logical ground there is only an analytic connection. The possibility of the latter requires no explanation, because it is possible according to the principle of contradiction. But the possibility of the connection between a real ground and its consequence poses a great problem.

One may observe here that according to proof-theoretic semantics, the necessary connection is provided by a categorical proof of B from A justifying the introduction of the implication.

Pleasing as this is, the intuitionistic view of harmony will not take one very far, once one accepts that non-monotonic reasoning is a viable form of reasoning. This is for, at least, the three following reasons:

1) In non-monotonic logics the relation between the specification of the meaning of a connective and inference patterns sanctioned by this meaning is very different from that relation in intuitionism. Some assignments of meanings to connectives do fix only trivial reasoning patterns for these connectives. This happens for instance with the implication in closed world reasoning. The meaning of this connective is specified in the following manner. If one has an implication $p \rightarrow q$, then one has MP as one component of the meaning of the implication. The second component is defined as "if one knows that $\neg p$, and also that there is no implication $r \rightarrow q$ for r different from p, then one can conclude $\neg q$." This is a perfectly formal specification of the meaning of the implication but what other inference patterns besides MP it sanctions is determined by the presence of other implications.

2) Dummett and Prawitz take it as axiomatic that a theory of meaning should be compositional. Here compositionality is meant in a very strong sense, namely that the lexical meanings that enter into the computation of the meaning of the sentence cannot be affected by context. An instance of this is natural deduction where each logical constant is defined by one or at most two introduction rules that are not dependent on context.

It is rather doubtful, however, whether natural language meaning is compositional in this strong sense. For instance, in closed world reasoning (involved in reasoning with exception-tolerant conditionals), given the premises $p \rightarrow q$ and $\neg p$, what consequence one can draw from this depends on the presence or absence of other premises with q as consequent. It follows that the meaning of implication as governed by negation as failure is determined by context.

Furthermore there exist perfectly natural reasoning systems such as closed world reasoning where rules are context-dependent, as we have seen in Chapter 2: whether AC is or is not allowed depends on context. Nevertheless, harmony is a general requirement on the relation between constitutive and regulative norms, so one also needs a notion of harmony for closed world reasoning. This topic will be taken up in Chapter 7.

3) The proof theoretic notion of validity of an argument Γ/φ is: any (canonical) proof of the premises Γ can be transformed into a proof of the conclusion φ. This forces monotonicity, since the proof of the premises Γ may actually prove the stronger $\Gamma \cup \Delta$ while still being capable of being transformed into a proof of φ. Thus, a stronger proof of the premises cannot invalidate the conclusion, unlike what is the case in closed world reasoning.

These three reasons suffice to show why the proof-theoretic analysis of harmony cannot be the final word. As will be explained in the next chapter it is even doubtful whether it is a correct description for intuitionism itself.

CHAPTER 7

This is the end?

The notion of harmony is appealing because it provides a convincing scheme for the justification of logical laws: reflect on the meaning of a connective and encapsulate this in an introduction rule, then add an elimination rule which respects harmony. Is this therefore the endpoint in the search for the normative character of logical laws?

Not really. It was remarked above in passing that the non-harmonious character of double negation elimination may be the consequence of an inadequate introduction rule for negation. Whether or not double negation elimination is derivable depends clearly on what proofs are possible in the system as a whole, and indeed on how the concept of proof is itself defined. For example, natural deduction systems have the property that the conclusion of a proof is always a single formula. In sequent calculi, this is not necessarily the case, and indeed it is possible to characterise negation entirely by rules involving a single negation only. Philosophically, this means that the vaunted locality of meaning-determination in natural deduction needs to be reconsidered: it is only by virtue of the overall structural constraints on proofs that the introduction rules can do their work in satisfying harmony.[1] In other words, if one considers a logical law to be unjustified, the blame cannot be put only on that rule itself, but must be distributed over the whole system of which it forms a part. In the present terminology this means that logical laws are justified only with respect to an entire logical form, and not because of either locality constraints or compositionality concerns. Harmony is an important idea, and it is a necessary ingredient for any theory of logical validity that does not define validity as truth preservation, but it cannot be formulated in the local manner that Dummett attempts.

1. Why is an extended notion of harmony necessary?

Even though the proof-theoretic notion of harmony was found wanting, there are still pressing conceptual reasons for assuming some form of harmony in the justification of logical laws. The need for harmony arises as soon as there are different ways to make a judgement, for example, via direct verification and via a more indirect proof. The Bolzano-Tarski definition of validity obscures this need by defining validity solely in terms of truth. Thus, if an indirect proof shows a statement to be false whereas the statement can be verified as true via direct inspection, the proof must be classified as an invalid argument, even though we may have been fully convinced of its validity beforehand. Theories of logic that assume truth and validity to be related but independent notions essentially require a notion of harmony.

We can find some clue as to what constitutes a good definition of harmony when we look again, but this time more carefully, at the intuitionistic example. Dummett and Prawitz strongly suggested that harmony is a concept concerned with the relation between rules. It is supposed to have a very local character, so that one can say of each particular pair

[1] As Michiel van Lambalgen pointed out to me, in the very constructive linear logic double negation elimination is totally unproblematic, due to the different nature of the structural rules.

of introduction and elimination rules whether it is harmonious or not. But we have seen that harmony or lack of harmony are properties of a proof system as a whole. A different definition of what a proof structure is could make double negation elimination harmonious. This shows that harmony is actually against the intention of its inventors a global notion. And it is not too far-fetched to say that it pertains more to an entire logical form than to inference rules.

There are other ways in which Dummett and Prawitz' analysis of harmony does not fully fit the intuitionistic case. It was remarked above that harmony forces monotonicity upon a logical system. While this is true for a given formal system, it does not seem to apply to intuitionistic provability generally. Here is a striking quote from Dummett to this effect:

> On any constructivist view of mathematics, on which its subject matter is our own mathematical activity, and meaning is given to our statements by reference to the methods of proof that we possess, this evolutionary process [i.e. the evolution of mathematics] must be understood more radically [than in Platonism], as entailing that the very meanings of our mathematical statements are always subject to shift. On the intuitionistic view, this evolution creates a special danger. If we look on the appeal to the full intuitionistic meaning of $\rightarrow$, in proving a statement of the form $A \rightarrow B$ as mediated by the invocation of a principle of the form $A \rightarrow$ there exists a proof of$A$, an advance in our apprehension of the available modes of proof may lead us to weaken such principles, because restrictions on the means whereby A could be proved which formerly seemed reasonable no longer appear so. When this happens, some proof, involving a conditional $A \rightarrow B$ that had formerly seemed acceptable, may be invalidated. Hence because of the peculiarities of the intuitionistic interpretation of $\rightarrow$, provability is not a stable property: we cannot think of an addition to our stock of methods of proof as merely allowing us to prove more than we could before, while all the proofs we had already given remain intact, since such an addition may lead to a rejection of certain earlier proofs. The intuitionistic interpretation of $\rightarrow$ does, indeed, give the notion of proof a self-reflexive or impredicative character, and to some degree weakens the conclusive and irreversible nature of mathematical results; mathematics becomes a subject whose results are fallible and liable to revision, like those of the other sciences [**10**, p. 401-2].

The phenomenon to which Dummett points can occur when one pushes the intuitionistic definition of implication to its limits. This definition says that there exists an effective construction which transforms any proof of the antecedent into a proof of the consequent. In Brouwer, a proof of an implication may take a 'transcendental' form[2] in that it proceeds via an investigation of all the ways the antecedent can be proved and showing that each of these can be transformed effectively into a proof of the consequent. However, if the proofs considered are not constrained to a single formal system, it may happen that a new proof construction of the antecedent comes up which has not been considered before and which cannot be transformed into a proof of the consequent. In this sense an intuitionistic proof of an implication can be defeasible. To drive the point home, one can say that taken to its extremes the underlying logic of intuitionistic mathematics is in principle non-monotonic.

2. On 'meaning as use'

In Prawitz and Dummett harmony as a property of rules constitutive of meaning is very much tied to the philosophical view of 'meaning as use'. There are several things to be said about this identification of which two are especially relevant to our purposes: i) whether

[2]The use of 'transcendental' here is due to Mark van Atten (unpublished ms., 2007).

all possible uses have been taken into account, and ii) whether meaning is at all exhausted by use.

As regards the first point, even if it is assumed that meaning is use, it is by no means clear that assertion and inference exhaust all possible uses. The combination of making an assumption and discharging it later seems to be an entirely different use[3] . Clearly, making assumptions and discharging them plays a very prominent role in natural deduction.

As regards the second point, one can make principled objections to the identification of meaning as use. The experimental data on the Wason selection task show clearly that someone’s use of an implication does not provide full access to her meaning of implication, although it is undoubtedly an important component. To use a simpler example here, if someone uses the conjunction elimination rule, this may indicate a correct usage of the logical conjunction or incorrect use of a conjunction with an implicative force as in Fillenbaum’s example [**13**] ‘do that and I’ll send you to your room’, which of course means ‘if you do this, I’ll send you to your room’. So what the use is and whether the use is correct can be identified only in the context of a larger interpretation. This entails that meaning is prior to use.

Another argument against the identification of meaning as use can be derived from closed world reasoning. In this type of reasoning, one can sometimes make inferential use of the implication as if it were a bi-implication, that is, affirmation of the consequent is sometimes justified. This usage, however, does not lead one to attribute a bi-conditional meaning to the implication; use is determined by other factors as well, in this particular case by the implementation of world-knowledge in the determination of the logical form. This observation is of course familiar from Gricean pragmatics, although no strict separation between semantics and pragmatics need be made here.

In fact, meaning is not the same as use even in Prawitz’ approach, except if some uses are more equal than others. If use is identified with ways of proving, not all proofs of a tautology can be regarded as constitutive of meaning. That any proof can be effectively transformed into a direct proof does not change the fact that the use of elimination rules draws on meanings already introduced and not the other way around. Strictly speaking, the use of introduction rules is identical to the meaning of logical constants, but the use of elimination rules is subsidiary to it. Observing this already opens up the theoretical possibility of different ways to relate direct and indirect proofs which will be taken upon in the following section.

Instead of trying to reduce meaning to use, making too much of the ‘effective method’ that provides the link, the different uses are better kept apart. In the end, as claimed earlier, it is only because of the general properties of the system that one can show something like the normalisation theorem to hold, which ultimately provides the license to reduce meaning to use. Even in this proof-theoretic context, it would be more precise to identify meaning with the introduction rules and use with the elimination rules; the first defining the rule-like meaning of the logical constants from which inferences by means of the elimination rules can be drawn. However, one can agree that assertion and inference from an assertion are important uses that have to be related by harmony, without assuming that all meaning is determined in this way.

---

[3]Kant makes this distinction of uses in his table of judgements, where he distinguishes between assertoric and problematic judgements [**27**, A70/B95].

3. Towards an extended notion of harmony

Stripped of all inessential features, the notion of harmony as introduced by Dummett and Prawitz boils down to the notion of harmony between the constitutive and regulative norms. For the introduction rules are used by Dummett and Prawitz as constitutive and the elimination rules as regulative and the latter must be justified on the basis of the former.

In fact, also in this case one needs a wider notion of constitutive norms, because the meaning of the connectives is constitutive not only through introduction rules but as well through general structural features of proofs. It is the entire package of structural proof features and introduction rules that form the constitutive norms to which the regulative norm must be adapted.

In the Brouwerian tradition of intuitionism such constitutive norms can even be conceived of as constitutive of part of our cognition, since after all proofs are just symbolic residues of mental constructions. This brings the notion of 'constitutive' closer to the Kantian sense.

4. Formal definition of extended notion of harmony

Defining harmony more formally and more generally brings issues to the fore that are somewhat obscured in the Dummett/Prawitz treatment. We can define the constitutive norm as what was previously called logical form, consisting of syntax, semantics and definition of validity. A regulative norm can in this context be conceived of as an inference pattern, and what is of interest now is the relation between these two.

One would like to say that a purported regulative norm must be derivable from a given constitutive norm. But this means both that the inference pattern is given as a concrete object and that there is a concrete derivation of the inference pattern from the parameter setting inherent in the constitutive norm. It will be seen that this transition from constitutive to regulative norm is by no means trivial. First, let us reformulate the Dummett/Prawitz theory of harmony into the present vocabulary. The syntax is that of an ordinary first order language and the basic semantic entities are the canonical proofs and the satisfaction relation, which holds between canonical proofs and formulas, is completely determined by the introduction rules. The definition of validity of an argument pattern is that every canonical proof of the premises must be transformable to the canonical proof of the conclusion. A justification of the regulative norms, i.e. the inference patterns, then proceeds by effective proof transformation, as indicated earlier. For the moment one can abstain from asking how the regulative norms are given, but this question will be taken up later. It is of interest to observe here that this effective proof transformation takes place in a meta-theory in which one can manipulate proofs. Clearly this meta-theory must have its own norms. This highlights the fact that in going from constitutive to regulative norms one may have to appeal to other norms. Perhaps these are fairly obvious in the case of Prawitz' proof normalisation but that need not always be the case.

Another instance where the derivability of regulative from constitutive norms is not trivial is furnished by closed world reasoning. The constitutive norm there is the logical form associated with negation as failure, but by itself this hardly determines any regulative norms. This is because the regulative norms must be derived from the completion, which is determined by world-knowledge. It follows that the required harmony can be defined only with respect to explicitly given world-knowledge. So in a sense, computing world-knowledge into the regulative norm restores harmony in closed world reasoning. This example shows that the derivation of regulative from constitutive norms may have to be aided by world-knowledge.

We now turn to the issue of how the regulative norms are given by the constitutive norm. There are both logical and cognitive aspects to this issue. The latter will be taken up in the next chapter. The logical issues have to do with decidability and completeness. One would like to have a procedure that generates the regulative norms given the constitutive norms. There are various obstacles to the fulfilment of this desire. For instance, since predicate logic and many other logics are undecidable, the generation procedure cannot be thought of as follows: take an inference pattern and determine whether it is sanctioned by the constitutive norm or not. The best one can hope for is to come up with a small set of regulative norms sanctioned by the constitutive norms from which all other regulative norms are derivable. This is possible if the constitutive norm is a logical form that has a completeness theorem. But completeness theorems come with a price, namely they often need strong principles to prove them, as is the case, for instance, with classical predicate logic where one needs a weak form of the axiom of choice. Let us recall the general outline of the completeness proof: one shows that if from Γ one cannot derive A, the set $\Gamma \cup \{\neg A\}$ has a model. This last step uses a weak form of the axiom of choice. If this weak form is not available, then $\Gamma \cup \{\neg A\}$ does not have a model, i.e. Γ entails A in the sense of Bolzano-Tarski validity, even though A is not derivable. So the extension of Bolzano-Tarski validity also depends on the background set theory – i.e. on considerations external to the form of the arguments. Thus, the bridge from constitutive to regulative norm now leads through an infinitary principle, which makes the connection between the two arguably much less direct than is desirable.

Sometimes what argument patterns are valid depends on assumptions about the world, or possible worlds. A basic distinction can be made between allowing (i) infinite domains or (ii) on the contrary finite domains only. In case (i), the completeness theorem for first order logic holds, so validity can be captured by natural deduction. In case (ii), we have a notion of validity that could be called ‘finite-validity’: an argument is finite-valid if every finite model of the premises verifies the conclusion. A theorem due to Trahtenbrot [**48**] says that no completeness theorem for finite-validity is possible. Technically speaking, every proposed proof system for this notion of validity is either inconsistent or will undergenerate. Note that the usual natural deduction proof system is sound for finite-validity, and since it cannot be complete, it must under-generate: there are finite-valid arguments which cannot be shown to be so by natural deduction. This also means that the modal element in finite-validity (’the conclusion follows necessarily from the premises’) cannot be captured by a proof system. This is interesting in view of the fact that Prawitz [**34**] advocates proof systems precisely because they would allow a characterisation of the modal element in validity.

A different way to pose the above question is to what extent the constitutive norms fully specify the regulative norms. As has been observed, the answer may depend on a very different set of norms used to make the transition from the constitutive to regulative. There are several possibilities for under-specification. If one sets the parameters equal to closed world reasoning, it is still not determined what the valid inferences are, because that very much depends on the knowledge one has. Closed world reasoning gives only a meta-norm, as discussed earlier, which yields object-norms in concrete knowledge situations.

Finally, if the regulative norms are of the form ‘parameter setting $\Longrightarrow$ set of inferences’ one may be plagued by under-specification.[4] For instance Russian constructivists accept ‘Markov’s Principle’, which says that if a predicate P is decidable, then $\neg\neg\exists x P(x)$ implies $\exists x P(x)$. Western constructivists in general do not accept this. That is, if the parameters are set so as to force constructive reasoning (e.g. the meaning of $\exists x P(x)$ is fixed to be ‘one can exhibit an n and a proof of $P(n)$’), then the logical principles are not yet

[4]The material for this paragraph is taken from Troelstra and van Dalen [**50**].

fully forced. This happens only if the parameters are set in a mathematical manner (e.g. by providing a fully precise semantics), so that it is possible to give an exact soundness proof.

This brings us to the next form of under-specification: the '$\Longrightarrow$' in the norm must be justified by means of mathematics (which itself involves logic, etc.). The dispute about Markov's Principle can also be viewed as a dispute about what mathematics to allow in the proof of '$\Longrightarrow$'. 'Westerners' claim that 'Russians' use classical considerations in their justification of '$\Longrightarrow$', instead of purely intuitionistic arguments. A related issue comes up when discussing completeness instead of soundness, for instance a meta-norm of the form 'parameter setting $\Longrightarrow$ all valid inferences can be derived from the finite set of inferences I'. As observed earlier, here '$\Longrightarrow$' may use some form of higher mathematics (e.g. a weak form of the axiom of choice for the classical completeness theorem), and disputes about the validity of these mathematical principles lead to some form of under-specification.

CHAPTER 8

Cognitive considerations: competence and performance

1. A new representational format is called for

This chapter discusses briefly what the preceding considerations mean for the cognitive substrate of human reasoning. By claiming a prominent role for logic in cognition (for instance, in the specification of what it is involved in ‘going beyond the information given’, the position adopted in this thesis clearly goes somewhat against recent trends in cognitive science. It is one thing, however, to claim an important role for logic, and quite another to explain how logic can do this. In this area, there are more questions than answers.

One route into a cognitive study of logical reasoning is via Chomsky’s distinction between competence and performance [**8**]. In linguistics, competence embraces a speaker’s implicit knowledge of his language, taken to consist of the lexicon and the syntactic rules. The sentences produced by the speaker are not part of his competence, but part of performance, the actual production of utterances. This distinction is meant to emphasise two things: the indefinitely large collection of sentences that the speaker can produce (or comprehend) are not part of his competence, which is assumed to be bounded; and performance of a real speaker can differ from that of an ideal speaker possessing the same competence because the former has, unlike the latter, cognitive limitations, such as bounds on the number of embedded constructions. These concepts are highly theory-laden, that is, they assume a very specific theory of language, but they will serve here as a rough guide.

Traditional psychologists of reasoning have assumed that the basic representational format for logic is either as rules (like for example MP) or as mental models and operations thereon. The former view is known as ‘mental logic’ (Rips [**37**]). It assumes that some inference rules (always supposed to be valid classically) have permanent representations, whereas others (for example MT) have to be computed on each occasion of use. This assumption would explain why MT is harder (i.e. has lower rate of endorsement) than MP. On this view, logical competence consists in the possession of the simple inference rules which have permanent representations, and of the correct ways of combining these in proofs of more complex rules. Performance can differ from competence especially in the combination of inference rules; some people are simply unable to combine the primitive rules for implication and negation into a proof of MT.

The contrary view, ‘mental models’ (Johnson-Laird [**25**]) maintains that inference proceeds by constructing models of the premises, where models are possibilities[1] allowed by the premises and reading off conclusions from these models; ideally, the output is true in all models of the premises. Competence consists in the ability to construct mental models, but performance may differ from competence due to memory limitations. If the premises of an argument have several different models (in practice: more than 2), it may be difficult to construct them all and check the truth of the purported conclusion in each of them. This is used to explain the highly variable degrees of difficulty in syllogistic inference.

[1]In the propositional case these possibilities can roughly be thought of as lines in a truth table; ‘mental models’ is thoroughly wedded to classical logic.

The notions of competence and performance that follow from the considerations in the preceding chapters must necessarily be very different. One reason is that the earlier approaches rely fully on classical logic. As a consequence, these approaches at best model reasoning from a fixed interpretation; they have nothing to say about the preceding stage, the reasoning to an interpretation whose goal is to impose a logical form allowing one to extract information from the data. A classical logical form is only one of the possible interpretations that reasoners might arrive at.

2. The varieties of competence

This distinction in two forms of reasoning involved in any inference task means that reasoning competence, and likewise performance, must have several components. Part of the competence consists in the ability to impose some logical form on the reasoning problem, although competence by itself does not dictate which. This component of competence is claimed to be universal in the sense that it is a feature of fundamental cognitive processes: imposition of logical form is necessary to go beyond the information given. Thus, the subject who imposes a classical logical form on the suppression task is neither more nor less competent than the subject imposing the closed world assumption. However, the competent subject must realise that inference can take place only after essential elements in the logical form have been fixed. As another example, a reasoner attempting a syllogistic task with a diagrammatic solution strategy must appreciate that she needs, besides a representation of predicates as circles (say) and a representation of a premise as a relation between circles, also a principled way of combining diagrams for the premises.[2] It is this integrative aspect of logical form that is usually most important in reasoning tasks, and hence in reasoning competence.

In the view just outlined, the basic representational format is neither ‘rules’ nor ‘models’, but something else entirely, logical form. This is a much more complex structure than either ‘rules’ or ‘models’. Whereas the latter are basically small finite sets, a logical form has a recursive structure: both the formal language and the notion of satisfaction are recursively defined. But the apparent simplicity of the traditional approaches is largely illusory, since the process which leads from reasoning task to representation is left out of consideration; these approaches concentrate on the product, i.e. the outcome of the process only. The process as it is captured in logical form is in a sense infinite, while the product (here inference rule, or model) is finite. So there is no way to reduce an infinite process to a finite product. From a finite sentence one cannot determine, for instance, what the formation rules of the sentence are as well as the semantics, definition of validity and so on. In order to do that, one needs to assume the sentence to be true, for which a definition of truth is already required. This topic will be returned to when discussing what information an experimenter can hope to get from reasoning tasks. Before leaving this aspect of competence, however, its high degree of idealisation must be noted. In fact the situation is rather analogous to natural language, where there are cognitive constraints on, for instance, the depth of embeddings; similarly one expects here only a very limited actual grasp of the recursion involved in logical form. What this means for normativity will be explored in the final chapter.

Parallel to this part of competence there is a notion of performance: the actual imposition of logical form. The experimental data show that subjects can experience considerable difficulties in doing this consistently. For instance, in the selection task it has been observed that subjects change the direction of the implication in the rule at issue according to whether they focus attention on the visible face, or on the invisible back of the cards.

[2] See Stenning and van Lambalgen [**44**, Chapter 10].

Thus, in these cases the logical form assigned to the selection task is not stable throughout the reasoning process. If this results in inconsistencies, one is definitely concerned with what can be called a performance error. It could be said that (this part of) competence embodies the norm that the various elements in the logical form must be chosen so as to yield a consistent form. However, one should not conclude from this that the same logical form must be maintained throughout the reasoning process, if only because the reasoner may start with an inappropriate logical form which he cannot but revise later. An example of this phenomenon was discussed at the beginning of chapter 5.

A second component of competence consists in having a method for finding inferences useful for the task at hand, which are sound for the logical form fixed in the first step. Whether or not this part is called upon very much depends on the difficulty of the task. In the selection task it rarely happens. In the suppression task it is the second component that is most prominent, although one may induce severe difficulties with the first component by reformulating the conditionals in a classical manner, using disjunction and negation: ‘she doesn’t have an essay or she studies late in the library’. This reformulation makes it very difficult for the subjects to assign a logical form.[3] In the syllogistic task it is again the second component that is most prominent. But what is important is that subjects can have the first component of competence without necessarily having the second component. It depends very much on how easy it is to go from the first to the second component.

Taking competence to consist in having a method for finding inferences implies that most of the inference patterns that go with a particular logical form are not part of the competence, although some may be. Applying the method, sanctioned by competence, to find a useful inference is again part of performance. The previous chapter contained some theoretical considerations on this aspect of competence and performance that will be briefly summarised here. The main observation was that going from competence to performance is itself a process of inference which may be more or less laborious depending on the logical form. For this theoretical reason alone one would not expect that, in general, performance comes anywhere near mirroring competence, in the sense that the inferences sanctioned (and the fallacies forbidden) by the chosen logical form are obvious to the reasoner. The reasoner may need ingenuity to come up with the desired inferences, or develop a heuristic based on the logical form, as some subjects succeed in doing when practising the syllogistic task.

It is only in the case of closed world reasoning that one would expect inferencing itself (as opposed to the imposition of logical form) to be relatively easy, because the logical form embodies the claim that the conclusions that follow from a given set of premises can be read off from a minimal model of these premises; and these models can be computed rapidly and automatically (Stenning and van Lambalgen [**44**, Chapter 8]). If the claims of van Lambalgen and Hamm [**52**] are correct, this fact is responsible for the automaticity of inferences in natural language comprehension.

While it is more or less clear how performance is related to competence in a theoretical sense, in practice it is often entirely unclear what to conclude about competence from observed performance. We have seen several examples of this in the preceding chapters. A subject who does not endorse AC may be motivated by classical logic, or by closed world reasoning applied to a hidden alternative conditional premise. A subject who concludes that the rule in the selection task ‘if there is a vowel on one side of the card, there is an even number on the other side’ is true, even though she has just turned the 7 card to find A, may have a notion of truth allowing exceptions, or may simply be confused. Another subject in the same task may take the A card to be sufficient to decide between truth and

[3] This observation is due to Henrik Nordmark.

falsity of the rule, because for her '$p \to q$ is false' means '$p \to \neg q$ is true'; or the subject may read the rule as 'if there is a vowel on the visible face, there is an even number on the invisible back'. These examples can be multiplied. There are so many parameters in the logical form to be estimated, that the data set has to be very large to conclude anything about competence from performance.

CHAPTER 9

Back to Normativity

In what sense is logic normative, then? The thesis started out (Chapters 1 and 2) by presenting some examples of logicians, psychologists and philosophers maintaining that logical laws (usually identified with classical logic) are normative *tout court*. We have seen that for actual human reasoning this implies that most people continually go against these norms and must thus be deemed to be irrational. This view has been seriously questioned by showing that human reasoning can often be satisfactorily modelled used alternative logics such as closed world reasoning (Chapters 2 and 3). Of course, earlier in the last century the primacy of classical logic had already been challenged by intuitionistic mathematicians, who claimed that the law of excluded middle is inadequately supported, if not outright false. In the meantime logics have continued to proliferate.

In the face of all this variety the question becomes pertinent of whether one can ascribe normative force to any set of logical laws. The answer explored in the thesis involves making a distinction between constitutive and regulative norms following Kant. We have seen that no experimental reasoning task speaks for itself; a task needs interpretation before it can be made sense of. In other words, the task is not given, but needs to be constituted or synthesised (Chapters 4 and 5). The constitutive norms are, therefore, the laws of the process of synthesis. In this scheme, analysis corresponds to the regulative norms, a kind of inference one can make on the basis of the interpretation assigned. What logicians earlier called normative force of logical laws can actually be ascribed only to what we here call regulative norms. It is only these regulative norms that justify the use of the notion of fallacy (Chapter 4). Since evaluation cannot but follow interpretation, it has been argued here that the regulative norms are always relative to the constitutive norms.

Normativity thus presents a Janus-face to us. On the one hand, logical laws are inescapable and a priori, on the other hand, they are revisable. Their inescapability they have in common with earlier conceptions of logical laws. Here, however, it simply means that one is bound to reason according to some set of logical laws, namely the logical laws constituting the interpretation that one constructs. But unlike traditional conceptions of logic, one must conceive of logical laws as revisable. The reason for this is that the assignment of a logical form always serves a purpose, namely, extraction of information that is deemed relevant to the task at hand (Chapter 2). Whenever the logical form is judged to be useless or at least not sufficiently effective for that purpose, it is liable to be changed. For example, one could adopt a closed world understanding of a reasoning task, to be replaced later by a classical understanding (Chapter 8). Such shifts are very likely to occur in one's attempt to solve a reasoning task. However, they should not be seen as instances of irrational behavior. Indeed there is no external norm to dictate a particular interpretation, when in principle more than one can be constructed[1].

1. The ideality of reasoning

Generally speaking, norms have an ideal character. Consider as an example the norm 'thou shall not kill'. This norm is ideal in two senses. First, one may subscribe to the universal character of this norm but hesitate about whether a particular act should be classified as an act of killing. In other words, a norm applies to idealised situations in which the acts forbidden are assumed to have exact definitions. In this they are no different from the laws of physics (Cartwright [**7**]). Norms are also ideal in a second sense, in that they are open to exceptions, not all of which have to be, or indeed can be, specified in advance.

These considerations also apply here. As we have seen in Chapter 8, reasoning competence, i.e. the set of constitutive norms, has a highly idealised character. This is because reasoning is a constructive process that starts from but goes much beyond the information given. However, apart from the ideality incurred by the necessity of imposing a logical form, the logical form itself also introduces many ideal elements, as seen in Chapter 8. There we concluded that the gap between competence and performance will generally be considerable. In other words, the constitutive norms are so highly idealised that conformity with these norms is not to be expected.

But what exactly does this ideal character of the norms entail for the experimenter? By taking logic to be normative, the analyst could once have the certainty of possessing the standards according to which human rationality can be measured. This means in our terms, that by using logic as a norm, the experimenter could dispense with the inference from performance to competence, and could restrict herself to comparing performance with the given norm. If the division of norms into constitutive and regulative norms proposed here is correct, this route is no longer open to the experimenter. The question, thus, arises: how may the experimenter get a grasp of the competence model that underlies the subjects' performance in a given task? In other words, what are the consequences of redefining normativity in the way it has been proposed here?

Inference to a competence model will always be accompanied by great uncertainty. But there are in principle two strategies opened to the experimenter. One is to divide a reasoning experiment in two parts; the first part designed to obtain information about the subjects' competence model, and the second part testing performance against the hypothesised competence model. The other strategy is to argue for a competence model from general cognitive considerations, as is done, for instance, in Stenning and van Lambalgen [**44**], where it is argued that much logical reasoning can actually be traced back to the human planning mechanism. Nonetheless, how to assess whether a performance error has been committed, remains an acute question. To obtain certainty would require a God's eye view of at least all possible competence models. In practice, therefore, performance errors function as some kind of rest category of events that are not otherwise explainable. In this,

[1]One significant question that has hardly been touched upon in this thesis concerns the notion of truth. The deflationary theory of truth holds that asserting 'it is true that p' is equivalent to p. Thus, seemingly obviating the need for more contentful theories of truth such as the correspondence theory. According to the deflationary theory, truth is a monadic property, albeit a trivial one. The notion of logical form adopted here embodies a very contentful theory of truth. Each logical form comes with its own notion of truth and, thus, what assertion means is completely relative to the constitutive norms which govern the context in which the assertion is made. This seems to be the very opposite of a deflationary theory of truth, unless there is a notion of truth prior to logical form. But this is a topic for another thesis.

however, the psychology of reasoning does not differ from other branches of science. After all, the experimenter has not other option than to impose her own logical form.

Another aspect of the ideal character of norms has come to the fore in the discussion of harmony in chapter 7. Harmony is a relation that ideally obtains between regulative and constitutive norms. The regulative norms must be fully justified on the basis of the constitutive norms only. Chapter 7 has observed, however, that this justification itself involves a reasoning process governed by norms, which may not be universally shared. Also, and more basic, the derivation of the regulative norms may themselves be complex and hardly attainable for the ordinary subject in a reasoning task. This happens, for instance, in the Wason selection task. Harmony, therefore, functions as a regulative ideal in the Kantian sense: one always strives for it, but it is unlikely to be attained.

Bibliography

[1] A. Athanasiadou and R. Dirven. Typology of *if*-clauses. In E. Casad, editor, *Cognitive linguistics in the redwoods*, pages 609–654. Mouton De Gruyter, 1995.

[2] A. Athanasiadou and R. Dirven. *On conditionals again*. John Benjamins, Amsterdam, 1997.

[3] J. Barwise and J. Etchemendy. *The Language of First-Order Logic*. CSLI Publications, Stanford, CA, 1993.

[4] L.E.J. Brouwer. *Brouwer's Cambridge lectures on intuitionism*. Cambridge University Press, Cambridge, 1981.

[5] J.S. Bruner. *Beyond the information given*. W.W. Norton & Co., New York, 1973.

[6] R.M.J. Byrne. Suppressing valid inferences with conditionals. *Cognition*, 31:61–83, 1989.

[7] N. Cartwright. *How the laws of physics lie*. Clarendon Press, Oxford, 1983.

[8] N. Chomsky. *Aspects of the theory of syntax*. MIT Press, Cambridge, MA, 1965.

[9] I.M. Copi. *Symbolic Logic*. MacMillan, New York, 1954.

[10] M. Dummett. *Elements of intuitionism*. Oxford Logic Guides. Oxford University Press, Oxford, 1977.

[11] M. Dummett. *The logical basis of metaphysics*. Duckworth, London, 1991.

[12] J. Etchemendy. *The concept of logical consequence*. Harvard University Press, Cambridge, MA, 1990.

[13] S.I. Fillenbaum. How to do some things with if. In Cotton and Klatzky, editors, *Semantic functions in cognition*. Lawrence Erlbaum Associates, 1978.

[14] A. Fisher. *The logic of real arguments*. Cambridge University Press, Cambridge, 2004.

[15] G. Frege. *The foundations of arithmetic*. Translated from the German by J.L. Austin. Oxford University Press, Oxford, 1950.

[16] G. Gebauer and D. Laming. Rational choices in Wason's selection task. *Psychological Research*, 60:284–293, 1997.

[17] M. C. Geis and A. M. Zwicky. On invited inferences. *Linguistic Enquiry*, 2:561–566, 1971.

[18] G. Gentzen. Investigations into logical deduction. In M.E. Szabo, editor, *The collected papers of Gerhard Gentzen*, Studies in logic and the foundations of mathematics. North-Holland Publishing Company, Amsterdam, 1969.

[19] P Hagoort and J. van Berkum. Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481):801–811, May 29 2007. DOI: 10.1098/rstb.2007.2089.

[20] A. Heyting. *Intuitionism: an introduction*. Studies in logic and the foundations of mathematics. North-Holland Publishing Company, Amsterdam, 1956.

[21] D Hilbert. The foundations of mathematics. In J. van Heijenoort, editor, *From Frege to Goedel*. Harvard University Press, Cambridge, 1967. Article originally appeared in 1927.

[22] D Hilbert. On the infinite. In J. van Heijenoort, editor, *From Frege to Goedel*. Harvard University Press, Cambridge, 1967. Article originally appeared in 1925.

[23] K.J.J. Hintikka. *Logic, language games and information*. Clarendon Press, Oxford, 1973.

[24] P. N. Johnson-Laird and R.M. Byrne. *Deduction*. Lawrence Erlbaum Associates, Hove, Sussex., 1991.

[25] P.N. Johnson-Laird. *Mental models*. Cambridge University Press, 1983.

[26] I. Kant. *Prolegomena zu einer jeden kuenftigen Metaphysik, die als Wissenschaft wird auftreten koennen*, volume IV of *Kant: Gesammelte Schriften*. De Gruyter, Berlin, 1968.

[27] I. Kant. *Critique of pure reason; translated from the GErman by Paul Guyer and ALlen W. Wood*. The Cambridge edition of the works of Immanuel Kant. Cambridge University Press, Cambridge, 1998.

[28] B. Longuenesse. *Kant and the capacity to judge*. Princeton University Press, 1998.

[29] E. Mendelson. *Introduction to Mathematical Logic*, volume New York. Van Nostrand, 1979.

[30] M. Oaksford and N. Chater. Probabilities and pragmatics in conditional inference: suppression and order effects. In D. Hardman and L. Macchi, editors, *Thinking: psychological perspectives on reasoning, judgment and decision making*, chapter 6, pages 95–122. John Wiley & Sons, Chichester, 2003.

[31] J. Piaget. *Logic and psychology*. Manchester University Press, Manchester, 1953.

[32] C.J. Posy. Brouwer's constructivism. *Synthese*, 27(1–2):125–159, 1974.

[33] D. Prawitz. *Natural deduction. A proof theoretic study*. Almquist & Wiksell, Stockholm, 1965.

[34] D. Prawitz. Proof and consequence. In S. Shapiro, editor, *The Oxford handbook of philosophy of mathematics and logic*. Oxford University Press, 2005.

[35] W.V.O. Quine. *From a logical point of view*. Harper & Row, New York, 1961.

[36] J. Rawls. Two concepts of rules. *The Philosophical Review*, 64(1):3–32, 1955.

[37] L.J. Rips. *The psychology of proof*. The M.I.T. Press, Cambridge, MA, 1994.

[38] B. Russell. *Our Knowledge of the External World (Lecture IV)*. Allen and Unwin, London, 1914.

[39] M. Sainsbury. *Logical form: an introduction to philosophical logic*. Blackwell, Oxford, 1991.

[40] J.R. Searle. *Speech acts*. Cambridge University Press, Cambridge, 1970.

[41] K. Stenning and M. van Lambalgen. Semantics as a foundation for psychology. *Journal of Logic, Language, and Information*, 10(3):273–317, 2001.

[42] K. Stenning and M. van Lambalgen. A little logic goes a long way: basing experiment on semantic theory in the cognitive science of conditional reasoning. *Cognitive Science*, 28(4):481–530, 2004.

[43] K. Stenning and M. van Lambalgen. Semantic interpretation as reasoning in nonmonotonic logic: the real meaning of the suppression task. *Cognitive Science*, 29(6):919–960, 2005.

[44] K. Stenning and M. van Lambalgen. *Human reasoning and cognitive science*. MIT University Press, Cambridge, MA., 2007.

[45] P.F. Strawson. *Introduction to logical theory*. Methuen, London, 1952.

[46] P.C. Suppes. *Introduction to Logic*. Van Nostrand, New York, 1957. Republished in 1999.

[47] A. Tarski. *Introduction to logic and the methodology of the exact sciences*. Oxford Logic Guides. Oxford University Press, Oxford, 1994. First published in 1941.

[48] B. A. Trahtenbrot. The impossibility of an algorithm for the decision problem for finite domains. *Doklady Academii Nauk SSSR*, 70:569–572, 1950.

[49] A.S. Troelstra and H. Schwichtenberg. *Basic proof theory*, volume 43 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, Cambridge, 1996.

[50] A.S. Troelstra and D. van Dalen. *Constructivism in mathematics. An Introduction*. Elsevier, Amsterdam, 1988. Studies in Logic Vols. 121 and 123.

[51] F.H. van Eemeren and R. Grootendorst. *A systematic theory of argumentation : the pragma-dialectical approach*. Cambridge University Press, Cambridge, 2004.

[52] M. van Lambalgen and F. Hamm. *The proper treatment of events*. Blackwell, Oxford and Boston, 2004.

[53] P. C. Wason. Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20:273–281, 1968.

[54] P.C. Wason. Problem solving. In R.L. Gregory, editor, *The Oxford companion to the mind*. Oxford University Press, Oxford, 1987.

[55] R.A. Wilson and F. Keil (eds). *The MIT Encyclopedia of the cognitive sciences*. MIT Press, Cambridge, MA, 1999.