Chapter 1
Formalization

Sven Ove Hansson

Abstract This introduction to formal philosophy has its focus on the basic methodology of formalization: the selection of concepts for formalization, appropriate splittings and merges of concepts to be formalized, the idealization that is necessary prior to formalization, the identification of variables and their domains, and the construction of a formal language. Other topics covered in this chapter are the advantages and pitfalls of formal philosophy, the relationships between formal models and that which they represent, and the use of non-logical models in philosophy.

1.1 Introduction

Few issues in philosophical style and methodology are as controversial among philosophers as formalization. Some philosophers are anti-formalists who consider texts making use of logical or mathematical notation as non-philosophical and not worth reading. Others are pan-formalists who consider non-formal treatments as – at best – useful preparations for the real work to be done in a formal language. But discussions on the pros and cons of formalization are more common at the coffee tables of philosophy departments than in scholarly books and journal articles. That is unfortunate since formalization has important methodological issues in need of systematic treatment.

This chapter is devoted to the use of formal methods in philosophy. It has a (non-exclusive) emphasis on logic which is the most commonly used formal language in philosophical investigations. We will have a close look at what formal logic is (Sect. 1.2) and how it can contribute to philosophical clarification (Sect. 1.3), the process that takes us from natural to logical language (Sect. 1.4), the construction of a logical language (Sect. 1.5), some philosophical uses of logical inference

S. O. Hansson (✉)
Division of Philosophy, Royal Institute of Technology (KTH), Stockholm, Sweden
e-mail: soh@kth.se

S. O. Hansson, V. F. Hendricks (eds.), *Introduction to Formal Philosophy*, Springer Undergraduate Texts in Philosophy, https://doi.org/10.1007/978-3-319-77434-3_1

(Sect. 1.6), and the philosophical use of non-logical formal models (Sect. 1.7). Finally, we will summarize some of the dangers and difficulties in the philosophical use of formal methods (Sect. 1.8).

1.2 Formal Logic as a Tool for Philosophy

Formal philosophy began with logic, and logic is still its dominating formal language. A good case can be made for increased use of non-logical formal methods, but in a general exposition of formal philosophy, logic is still the best starting-point.

1.2.1 The Origins of Logic

Logic is concerned with how we draw conclusions. Its systematic study begins with the observation that some inferences fall into general patterns. These patterns are characterized by being insensitive to the meaning of certain elements of that which we say or think, and even unaffected by the uniform substitution of these elements. Following Gottfried Wilhelm Leibniz (1646–1716), we can use the term "formal arguments" for arguments in which "the form of reasoning has been demonstrated in advance so that one is sure of not going wrong with it" [49, p. 479].[1] Consider the following argument:

Rich men are condescending.
Therefore: Non-condescending men are not rich.

The changeable elements here are of course "rich men" and "condescending men". We will call them variables. The example exhibits three important features of variables in a logical argument. First, the validity of an argument is unaffected by vagueness in its variables. In most other contexts, the use of vague terms makes it difficult to determine whether that which is said is valid or not. Thus, the sentence "He is rich" is vague because the term "rich" is vague, and for a similar reason so is the sentence "He has condescending manners". But the inclusion of both these vague terms into the above argument does not affect its validity.

Secondly, the validity of an argument does not depend on whether that which is said about the variables is true or false. Suppose that you meet the richest man in the world and he turns out to be a friendly and respectful person. Then the premise of the argument is not true, but the argument is still valid, i.e. it is still true that the conclusion follows from the premise.

[1] "... des argumens en forme; parce que leur forme de raisonner a esté prédemontrée, en sorte qu'on est seur de ne s'y point tromper" [48, 478–479].

Thirdly, we can freely substitute the variables for something else, if we do so uniformly. By uniformity is meant that all instances of a variable are substituted by the same new element. We can for instance make the following substitution in the above argument:

Baroque music is beautiful.
Therefore: Non-beautiful music is not Baroque music.

We know that this argument is valid since the previous one is valid. They are instances of the same argument form. When analyzing an inference, it is useful to express it in such an argument form.

The above examples represent an argument form with one premise and one conclusion. In Aristotle's (384–322 BCE) logic, such arguments are called conversion rules. Aristotle referred for instance to the argument form exemplified by the conclusion from "No pleasure is good" to "No good is a pleasure" ([2], I:ii, 25a). However, the major focus in Aristotelian logic was on arguments with two premises and one conclusion, called syllogisms. The following is an example of a syllogism:

All logicians are philosophers.
Some logicians are cacographers.
Therefore: Some philosophers are cacographers.

The validity of this syllogism is not disturbed by the vagueness of the terms "logician" and "philosopher". Even more importantly, to confirm the validity of this argument one need not know what a cacographer is – or for that matter what a philosopher or a logician is.

Archimedes (c.287-c.212 BCE) is reported to have said: "Give me a place to stand on, and I will move the Earth" [16]. For a lever to work properly, we need a rigid and reliable pivot. Similarly in logic, in order for some terms, namely the variables, to be flexible in meaning and indeed exchangeable, we need other terms that provide a rigid and immutable platform on which the movements and exchanges of variables can take place. The terms that have this function are called *logical constants*. In the above examples, "all", "some", and "not" have the role of logical constants. Syllogistic logic, which held sway from Aristotle's time until the late nineteenth century, was devoted to these three logical constants and the argument forms that could be constructed with them. But there were also three parallel traditions in logic that employed other logical constants.

One of these was sentential logic, the logic of sentences, first developed by Chrysippus (c.279-c.206 BCE) and other Stoics. In sentential logic, the variables are sentences or propositions, rather than parts of sentences as in syllogistic logic. Chrysippus accurately identified a proposition as "that which is capable of being denied or affirmed as it is in itself" [21, pp. 69–70]. The logical constants are words like "and", "or", "if", and "not". An argument in sentential logic can be as follows:

Either I laugh or you cry.
I do not laugh.
Therefore: You cry.

Here, the variables are the sentences "I laugh" and "You cry", and the logical constants are "not" and "and".[2] Sentential logic lived a marginal existence in the shadow of syllogistic logic but gained in importance through the work of George Boole (1815–1864) and others in the nineteenth century.

The second of these traditions was modal logic, the logic of necessity, possibility, and related concepts. Its two most important logical constants are "necessarily" and "possibly". The oldest texts on modal logic are by Aristotle himself. Just like sentential logic, modal logic was overshadowed by standard syllogistic logic. It was revived in the early twentieth century by C.I. Lewis (1883–1964).

The third tradition is somewhat more difficult to pinpoint. It has its origin in what Aristotle called the *topoi*, or topics. These were valid arguments in which the role of logical constants was played by a wider range of concepts. These include "good", "better", and "child". Studies of the topics continued through the ages, although usually with somewhat less precision than in the dominant logical pursuit, namely syllogistic logic [23]. The importance of such argumentation was emphasized by Leibniz when he wrote:

> "It should also be realized that there are *valid non-syllogistic inferences* which cannot be rigorously demonstrated in any syllogism unless the terms are changed a little, and this altering of the terms is the non-syllogistic inference. There are several of these, including arguments from the direct to the oblique – e.g. 'If Jesus Christ is God, then the mother of Jesus Christ is the mother of God'. And again, the argument-form which some good logicians have called relation-conversion, as illustrated by the inference: 'If David is the father of Solomon, then certainly Solomon is the son of David." ([48], p. 479; translation from [49], p. 479)

1.2.2 The "Newtonian" Revolution in Logic

These traditions in logic – studies of syllogisms as well as the other, subsidiary subject areas – had one important limitation in common: They were devoted to single argumentative steps. Actual argumentation usually proceeds by a whole series of steps. This restriction to single steps, taken one at time, turns out to be a serious limitation since some arguments cannot be fully understood unless one takes a more comprehensive approach. Clear examples of this can be found in mathematical reasoning. In his *Elements*, Euclid (fl.300 BCE) often introduced an assumption only in order to refute it. After making the assumption he presented a multi-step argument based on it. Many steps later he arrived at an inconsistent conclusion, based on which he inferred that the assumption was false ("reductio ad absurdum", reduction to absurdity) This is a type of argumentation that logicians had great difficulties in accounting for since they dealt with each step separately [45, p. 597].

In the middle of the nineteenth century, logic was still a particularistic discipline, dealing with small argumentative steps in isolation, and lacking a unifying theory

[2]This is the argument form later known as Modus tollendo ponens or the disjunctive syllogism [5].

for the various types of argumentative steps. We can compare its status to that of mechanics two hundred years earlier. Before Isaac Newton's (1642–1727) *Principia* (1687), there were two branches of mechanics: terrestrial mechanics that dealt with the movements of objects on earth and celestial mechanics that dealt with the movements of heavenly bodies. Newton managed to unite the two disciplines by providing a mathematical model that was sufficiently general to cover the movements of both earthly and heavenly bodies. His new framework covered not only single events, but also complex interactions among a large number of objects, such as the bodies of the solar system.

In 1879 Gottlob Frege (1848–1925) published his *Begriffsschrift* which did to logic what the *Principia* had done to mechanics [19]. Frege's major invention was a notation (quantifiers) that could express the logical constants "all" and "some" in a much more versatile manner, and made them easily combinable with sentential constants such as "and" and "or". His new framework was a general logical calculus lacking the limitation to small steps that was inherent in the Aristotelian system of syllogisms. Instead of considering just two premises it was now possible to consider any set of premises, however large. This made it possible to ask questions that did not even arise in the logic of syllogisms. For any given a set of premises, one could ask whether a particular conclusion follows from it. Sometimes that question could be answered affirmatively by providing a step-by-step proof. In other cases it could be answered negatively by showing that no combination of valid argumentative steps can lead to the conclusion. With Frege, logic took the giant leap from an atomistic study of the smallest parts of arguments to a holistic analysis of what can and cannot be inferred from given premises.

Frege's system was limited to the logical constants that had been studied for more than two millennia in syllogistic and sentential logic: "all", "some", "not", "and", "or", "if", and "if and only if". Including them all in one and the same system was a major achievement, not least since arguments using these logical constants cover a large part of mathematical reasoning. But for philosophy this was still not enough. In philosophical argumentation the structural properties of other terms than these have crucial roles. For instance, if we wish to scrutinize Kant's views on whether ought implies can, then we do not have much use for the logical principles governing words like "all" or "and". Instead, our focus will have to be on properties of the concepts expressed by the words "ought" and "can" [75]. In the twentieth and twenty-first centuries, philosophical logicians fully realized this, and developed logical systems in which the role of logical constants is played by terms representing a wide variety of notions such as "necessary", "possible", "know", "believe", "do", "try", "after", "permit", "decide", "will", "right", "good", "blameworthy", "duty", "better", "cause", "freedom", "vague", and a wealth of others. Many of these had been studied by logicians in previous centuries, as part of the modal or the topics tradition. However, after Frege they could be included in holistic systems of argumentation, rather than being used in rules referring to a single, isolated step of reasoning. Through all these extensions, formal logic has expanded its territory most substantially, and this expansion is still an on-going process. We can see it as the second step of the "Newtonian" revolution in logic, after the first step for which Gottlob Frege was himself responsible.

1.2.3 The Actual Truth or a Model of the Truth?

The remarkable achievements of Frege's system of logic inspired many philosophers, and some believed that logical analysis could now replace other, more uncertain methods used by philosophers. Bertrand Russell (1872–1970) maintained that "every philosophical problem, when it is subjected to the necessary analysis and purification, is found either to be not really philosophical at all, or else to be, in the sense in which we are using the word, logical" [71, p. 14]. He and many others believed that logic would make it possible to reach a more fundamental level of philosophical insight, thereby resolving philosophical problems that could not be solved in natural language due to its lack of precision.

It was soon discovered, however, that philosophers can disagree about a problem expressed in logical terms just as they can disagree about one expressed in natural language. Russell's own analysis of definite descriptions provides a clear example of this. By a definite description is meant one that applies to exactly one object. In English, definite descriptions are often expressed with the definite article "the" followed by a singular: "the teapot on the lowest shelf", "the current president of South Africa", etc. The problematic cases are those in which there is either no object or more than one object answering to the description. If I ask you to take out the teapot on the lowest shelf, you will have problems in following the instruction if there is either no teapot or two or more teapots on that shelf. The following standard example has been used in the discussion:

The king of France is wise.

According to Russell [69], this should be interpreted as follows in predicate logic, with K standing for "is the king of France" and W for "is wise":

$$(\exists x)(Kx \,\&\, (\forall y)(Ky \rightarrow x = y) \,\&\, Wx)$$

This can be paraphrased as follows: "There is ($\exists$) someone (x) who is king of France (K). Everyone ($\forall y$) who is king of France is identical to him. He is wise (W)." It follows directly from this analysis that (as long as France remains a republic) the quoted sentence is false.

In a criticism of Russell's account, P.F. Strawson (1919–2006) contended that if someone uttered the sentence "The king of France is wise", then the question whether that sentence was true or false "simply didn't arise, because there was no such person as the king of France" [76, p. 330]. In Strawson's view, our sentence can be formalized in the simple way

$$Wk$$

where W denotes "is wise" and k denotes "the king of France". According to Strawson, this sentence expressed a true statement when uttered in the reign of Louis XIV, and a false statement when pronounced in the reign of Louis XV. But when asserted during the time of a French republic it expresses no statement at all, and consequently the question whether it expresses a true or a false statement does not

even arise. Russell [70] disagreed and defended his original standpoint. The debate has continued since then [17].

This and many other examples show that merely translating a philosophical problem into logical language cannot be expected to solve it. Philosophical dispute can continue, now referring to the logical formulation. What logic can do, however, is to provide more precise statements of the problem and of alternative standpoints pertaining to its solution or dissolution. This, as we will see, can be an important enough achievement.

1.2.4 A Guarded Defence of Formalization

In a larger perspective, the rise of modern symbolic logic can be seen as part of a more general, long-term, trend: More and more scientific and scholarly disciplines have become dependant on mathematical modelling. Astronomy is the only empirical branch of learning that has been thoroughly mathematized ever since antiquity. Physics became gradually more and more mathematized from the late Middle Ages onwards, and chemistry since the late eighteenth century. But the great rush came in the twentieth century, when discipline after discipline adopted mathematical methods. One of the best examples is economics, which has gone from almost no use of mathematics to being dominated by theories expressed in mathematical language [14]. In the last few decades, formal models, in particular game theory, have had a strong and increasing influence throughout the social sciences. At the same time, the mathematization of the natural sciences has accelerated. Today, large parts of biology and the earth sciences, such as ecology, population genetics, and climatology, are thoroughly mathematized.

The reason why mathematical tools were adopted in these and many other areas is of course that they have proven efficient; they have improved the predictive and explanatory capacities of the disciplines in question. The increased role of formal methods in philosophy has a similar explanation: we have introduced formal tools in order to express problems more precisely and obtain solutions in new ways. But there is a caveat: The usefulness of formal tools is not quite as overwhelming in philosophy as in the empirical disciplines. The difference can be seen from a comparison between philosophy and early physics.

We usually think of mathematical physics as beginning with Galilei Galileo (1564–1642), but mathematical methods were used in physics already in the fourteenth century. When medieval physicists (the so-called calculatores) developed mathematical models of physical phenomena, they proceeded in much the same way as Euclidean geometers. A geometer used "pure thought" to determine the laws that govern lines, surfaces, and three-dimensional bodies. In much the same way, physicists used their intuition when attempting to find the laws that govern the movement of bodies. And importantly, intuition had a double role: Not only was the development of these mathematical models guided by intuition, it was also against intuition that they were tested. This was before the great revolution in

physics led by Galileo. Although Galileo used his intuition as a starting-point when developing mathematical models of physical phenomena, he went on to test these models against experiments and exact observations. Since our mechanical intuitions are rather consistently wrong, this reality check was necessary to correct errors in the previous models [54, 74, 83].

Today, this is the standard approach to mathematical models in the empirical sciences. Mathematical models are tested against measurements whose values are expected to correspond to the variables of these models. Obviously, this can only be done if accurate measurement methods are in place. Before the thermometer was invented (in the seventeenth century), physicists had no better means to assess theories about heat than to compare them with everyday experiences of heat and cold. Exact measurement of temperature was a necessary condition for developing accurate mathematical theories of heat (thermodynamics). Today, no physicist would argue in favour of a thermodynamic principle by referring to our vague everyday experiences of heat and cold.

This is a general pattern in science. Measurement is our bridge between theories and observations. Mathematics is the medium in which we can transport information across that bridge, a medium unsurpassed in its information-carrying capacity. Today the bridge of measurement is quite crowded, carrying loads of information back and forth that are used on one side for the improvement of theories, and on the other side for the construction of new experiments and observations.

As we saw, physics had access to the mathematical medium long before it learnt how to avail itself of the bridge. Unfortunately, philosophy is in a situation comparable to that of pre-Galilean physics: we have the mathematical medium, but we do not have the bridge of measurement. And this is not a deficiency that can easily be mended within the confines of philosophy as we conceive it today. Philosophers studying concepts such as knowledge, truth, goodness, and permission are operating with constructs of the human mind that do not necessarily have exact empirical correlates. Our situation can to some extent be compared to that of mathematicians, who have all of their foundations on the theoretical side. Their research can improve the theories that are used in empirical work, but the information received back from empirical investigations does not normally lead to corrections of the mathematics. Similarly, philosophy can sometimes be used to improve theories in other disciplines, and the exactness of formal philosophy is often needed to match the precision required in these disciplines. But at least in most philosophical subject areas, empirical observations cannot support or disprove a theoretical statement in the same clear-cut way as in the empirical sciences.

Therefore, the claims that can be made for formalization are weaker in philosophy than in the natural and social sciences. In philosophy, the major virtue of formalization is the same as that of idealization in informal languages: By isolating important aspects it helps to bring them to light. In philosophical discussions we usually deviate from the general-language meanings of key terms such as "knowledge" or "value", giving them meanings that are more streamlined and more accessible to exact definition. This does not necessarily mean that we have access to a true philosophical meaning that these concepts should be adjusted to. A

much more credible justification is that such simplifications are necessary in order to obtain the precision needed for philosophical analysis. However, this is a sail between Scylla and Charybdis (on the bridgeless waters just referred to). We have to deviate from general language in order to make a sufficiently precise analysis. But if we deviate so far as to lose contact with general-language meanings, then the rationale for the whole undertaking may well be lost. This precarious situation applies, of course, to formal and informal philosophy alike.

All this boils down to a rather guarded defence of formalization in philosophy. It is a language in which we can build more precise models of philosophical subject matter, and as we will see, there are philosophical topics for which this increased precision is indispensable. However, formalization is no panacea. Mistaken ideas can be as easily formalized as valid ones. But although formalization is no safe road to philosophical truth, it is one of the best tools that we have for expressing, criticizing, and improving philosophical standpoints. It is an obvious but important corollary of this line of defence that we should not expect to find a uniquely "correct" formal analysis of philosophical subject matter. Different formalizations may capture different properties of our concepts [33, 38, 39].

1.3 Formalization as Clarification

The use of formalization in philosophy is part of our general strivings for clarity and precision in philosophical discussions. In this sense, formalization is continuous with the development of specialized (non-formal) philosophical language. Since antiquity, philosophers have spent much effort on clarifying the central concepts of the discussions they have taken part in, and almost invariably such clarifications have led to new distinctions and opened up for the formulation of new standpoints and new questions. We find such linguistic analysis in Plato's Socratic dialogues, for instance the discussions on virtue in *Meno* and knowledge in *Theaetetus*. We also find it in ancient texts from other civilizations, for instance in writings in the Mohist tradition in China that in many ways anticipated modern developments in the philosophy of language [52, 53].

1.3.1 The Need for Clarity

Clarity is still a major criterion of philosophical quality. We need precise concepts in order to develop and criticize philosophical arguments, and therefore careful analysis and development of our own terminology is an essential part of modern philosophy. This type of work is also an important part of philosophy's contributions to other disciplines. In interdisciplinary co-operations, it is often the role of philosophers to work out precise definitions and distinctions [34]. The importance of precision has been pointed out by many of the great philosophers, for instance by

Aristotle and (arguably with some amount of rhetorical exaggeration) by Ludwig Wittgenstein (1889–1951):

> "Our discussion will be adequate if it has as much clearness as the subject-matter admits of, for precision is not to be sought for alike in all discussions, any more than in all the products of the crafts." (Aristotle, Nicomachean Ethics I:iii, 1094b [3])

> "Everything that can be said can be said clearly." (Wittgenstein, Tractatus logico-philosophicus 4.116 [85])

So why should we strive for clarity and exactness? To begin with, we do so in order to facilitate communication. In everyday life we appreciate exactness whenever information is important for us. When listening to my stories about what I have seen in the streets of Berlin you probably do not worry much about how accurately I describe the geographical relations between the different streets, but if I give you directions to your hotel you will expect me to be quite precise about such details. If someone tells you about the medicine her aunt took against arthritis you may even prefer not to hear all the details about dosage and the like, but if your doctor recommends you to take a drug you want her to be very clear about doses and timing. As philosophers we are professionally interested in issues and details that most people seldom worry about, and therefore we often strive for exactness and clarity in respects that are usually disregarded in other contexts.

In addition to facilitating communication, exactness also facilitates investigation. If it is unclear to you exactly what I have said, how can you verify or repudiate my statement? As noted by Karl Popper (1902–1994), a statement has to be precise in order to be accessible to falsification or corroboration [47, 66]. This applies, of course, not only to philosophy but to science in general. One of the major virtues of mathematical theories in the natural and social sciences is that they provide us with predictions that are precise enough for testing.

In philosophy, as well as other disciplines, we often have to extend our language in order to express new thoughts and talk about that which we have not spoken of before. This is taken for self-evident in most academic disciplines. No one would expect a natural language to contain beforehand all the terms and distinctions needed to express new developments in chemistry, mathematics, or economics. In philosophy as well, new terms have been introduced along with new ideas and concepts. "Supervene", "induction", "modality", "consequentialism", and "prioritarianism" are examples of this.

Unfortunately, though, some philosophers seem to have believed that philosophical insights are in some way hidden in the language (mostly their own mother tongue). They have attempted to do philosophy by looking for meanings or connotations that only a person with an accurate feeling for the finest nuances of the language can pick up. But very few insights of lasting or general philosophical interest have been obtained in that way. The so-called ordinary language philosophy was a cul-de-sac. In order to develop philosophical terminology, we need to carefully construct and delimit new distinctions that have no obvious counterparts in non-philosophical language, and assign terms to them.

Since antiquity onwards, philosophy and poetry have been each other's antithesis in terms of their approaches to language. This may seem paradoxical since philosophy and poetry are closely related in another important respect: They both deal with "big issues" such as existence, meaning, knowability, and morality. But the two pursuits deal with these issues in different ways – ways that are complementary rather than competing. These differences have large effects on their respective linguistic ideals. In poetry, elegance usually has precedence over precision. In philosophy the reverse is usually the case, as keenly pointed out by C.S. Peirce when advocating

> "... a suitable technical nomenclature, whose every term has a single definite meaning universally accepted among students of the subject, and whose vocables have no such sweetness or charms as might tempt loose writers to abuse them – which is a virtue of scientific nomenclature too little appreciated." [65, pp. 163–164]"

In poetry, and in belles lettres in general, disambiguation is no goal. To the contrary, ambiguity and imprecision are often necessary means to achieve the desired literary effect [46]. Philosophy does the very contrary: It tries to achieve as much precision as possible, even though its subject matter often makes this particularly difficult [77].

1.3.2 What is Exactness?

Clarity is a wider concept that exactness. In order for a statement to be clear it is not sufficient for it to be exact. It also has to be expressed in a way that makes it reasonably easy to understand. Something that is clear should, in Descartes' words, be "open to the attending mind"[3] ([15, p. 22], [20]). For our present purposes we can focus on the somewhat narrower concept of exactness. ("Exact" can be taken to be synonymous with "precise".) This is a concept with two clearly distinguishable meanings. The following examples serve to show the difference:

(a) The colour of that laser beam is green.
(b) The colour of that laser beam is yellowish green.
(c) The colour of that laser beam lies somewhere in the wavelength interval 495–570 nanometres.
(d) She is in the centre of Paris.
(e) She is close to Notre Dame.
(f) She is in one of the first six arrondissements of Paris.

When going from (a) to (b) we restrict the scope of colours. Fewer colours answer to the latter than the former description. If we instead go from (a) to (c), we do not reduce the number of possible colours, or at least we are not sure to do so since "green" corresponds approximately to the stated wavelength interval. However,

[3] "Claram voco illam, quae menti attenditi praesens et aperta est."

(c) is considerably less vague than (a) since we have in practice eliminated the borderline cases that might be classified as either green or not green. Both the move from (a) to (b) and that from (a) to (c) can be described as moves in the direction of exactness, but these are different types of exactness. We can describe (b) as more restricted than (a), and (c) as more definite than (a). Similarly, (e) is more restricted than (d) and (f) more definite than (d). Restrictedness and definiteness are the two major forms of exactness.

It turns out that philosophically speaking, not even exactness itself is a sufficiently exact concept! [47, 77, 84] This can be seen clearly if we ask the simple question which of (b) and (c) is the more exact statement. The best answer to that question is to refuse answering it, and instead distinguish between the two notions of exactness, restrictedness and definiteness.

In philosophy, both types of exactness are important, but lack of definiteness tends to be more detrimental than lack of restrictedness. We can for instance use a wide concept of "action" that includes omissions (refraining from acting) and various non-intentional behaviour. Such a wide concept may be impractical for some purposes, but if its boundaries are sharp enough it will not create communicative hurdles that we cannot deal with. A concept of action that lacks in definiteness will be much more problematic, in particular if the undetermined borderline cases are among those that we need to attend to. Needless to say, the importation of such indefiniteness into a formal model will make the latter just as loose and ill-defined as its informal counterpart, and perhaps even more dangerously so if its vagueness is obscured by the seemingly exact paraphernalia of a mathematical language.

1.3.3 Can Inexactness Be Described Exactly?

We have to be realistic. Using the tools of philosophical analysis, we can make our concepts more specific and, in particular, more definite. But this is one of the many human activities in which perfection is in practice unattainable. Even after considerable efforts, many of our concepts will remain imprecise. Furthermore, some of the concepts that we wish to include in our analysis may be "essentially inexact", i.e. inexactness is part of what they express, and therefore their meaning cannot be mirrored by a definition from which the vagueness has been removed [34]. The relational concept "near" may be a case in point. Any precise definition of that concept, for instance as "within a distance smaller than 5.3 km" can be accused of missing essential features of nearness, namely that it comes in degrees and that it is judged differently in different contexts. (For instance, 5.3 km is near if you are driving on the motorway, but not if you are travelling by foot on an arduous mountain trail.) The same applies to concepts such as "bald" and "tall".

In such cases, instead of a vagueness-resolving definition we may opt for a vagueness-preserving one. The question then arises: Can we provide a formal representation that preserves the vagueness? The most obvious way to do so would be to construct a model in which the concept in question comes in degrees. We

can for instance construct a model in which I am tall to the degree 0.45 and the late basketball player Manute Bol (who was 46 cm taller than me) tall to the degree 0.99. But if the notion of tallness is essentially imprecise, as we have supposed, then it cannot be captured by such exact numbers. Perhaps we should make the numbers less precise, and assign to me tallness to the degree 0.40–0.55? But then both the lower and the upper limit appear to be artificially precise. Perhaps we should replace each of them by something less precise, such as an interval? In this way we are caught in an infinite regress of dissolving boundaries that seems very difficult to stop. Arguments like these have led some philosophers to question whether vague concepts can at all be adequately represented in a formal language [72, 78].

But there is a way out, for which we have already prepared the ground. A model should not be expected to correspond exactly to that which it is a model of. All that we can expect is that some features of the model should be structurally interrelated in the same way as some important features of the original. The grass mats used in a model railway may consist of plastic, but they represent lawns and meadows, not plastic mats. Similarly, the exact numbers in our model of degrees of tallness do not represent precise degrees. Instead, they represent the vagueness of our intuitive concept of tallness. They do this remarkably better than a model with all-or-nothing tallness, but they do not correspond perfectly to the intuitive concept. This should not be a problem, once we have realized that our formal construction is a model of our intuitive notion, not the "real truth" behind it. "[W]e can have mathematical precision in the semantics without attributing it to the natural language being studied by making use of the logic as modelling picture" [12, p. 246].[4]

1.4 From Natural Language to Logical Representation

Any representation of a concept in logic or some other formal language is the outcome of a streamlining of the concept, a simplification for the sake of clarity, in other words an idealization. In this section we are first going to have a close look at philosophical idealization, and in particular the relationship between formal and informal idealizations. After that we will turn to some of the major problems that have to be solved in the process of formalization, or idealization into a formal language as it can also be called. Throughout this section, the examples will concern logical formalization although most of the principles discussed are also relevant for formalization into other formal languages.

[4] Cf. Williamson's [84, pp. 270–275] notion of a "variable margin model".

1.4.1 The Nature of Idealization

Formal models are ideals in the sense of "[s]omething existing only as a mental conception". (OED) To idealize in this sense means to perform a "deliberate simplifying of something complicated (a situation, a concept, etc.) with a view to achieving at least a partial understanding of that thing. It may involve a distortion of the original or it can simply mean a leaving aside of some components in a complex in order to focus the better on the remaining ones" [61, p. 248].

This sense of idealization must be distinguished from the more common sense of expressing a (too) high opinion of something. Formal models may or may not represent something as "perfect or supremely excellent in its kind". (OED) In (formal and informal) philosophy, both types of idealization are common. In particular, the concepts that we use when philosophizing on human behaviour tend to be both (1) idealizing−simplifying, i.e. they leave out many of the complexities of real life, and (2) idealizing−perfecting, usually by representing patterns that satisfy higher standards of rationality than what most humans live up to [37]. Since formal philosophy has its starting-points in informal philosophy, it tends to inherit both types of idealization.

The reason why we idealize−simplify is that philosophical subject-matter is typically so complex that an attempt to cover all aspects will entangle the model to the point of making it useless. A reasonably simple model has to leave out some philosophically relevant features. For a simple example of this we can consider philosophical usage of the term "better". In ordinary language, "A is better than B" and "B is worse than A" are not always exchangeable. It would for instance be strange to say: "Zubin Mehta and Daniel Nazareth are two excellent conductors. Nazareth is worse than Mehta." Given the first sentence, the second should be: "Mehta is better than Nazareth." Generally speaking, we only use "worse" when emphasizing the badness of the lower-ranked alternative ([25, p. 13]; [80, p. 10]; [11, p. 244]). There may also be other psychological or linguistic asymmetries between betterness and worseness [79, p. 1060]. However, a long-standing philosophical tradition persists in not making this distinction in regimented philosophical language [7, p. 97]. The reason for this is that the distinction does not seem to have enough philosophical significance to be worth the complications that it would give rise to. The logic of preference adheres to this tradition from informal philosophy, and $A > B$ is taken to represent "B is worse than A" as well as "A is better than B".

Idealization−simplifying − be it formal or informal − always involves deviations from that which we model. Therefore, counter-arguments can always be made against an idealized account of philosophical subject matter. It is for instance easy to find examples in which betterness and worseness are not interdefinable in the way described above. Such deviations will always have to be judged in relation to the purpose of the model and how it is used. Does the deviation show that an important aspect of the subject matter has been "lost in idealization"? If it does, then we have to consider how much we would lose in simplicity by including it. Sometimes, the best strategy is to replace the idealization by a richer account. On other occasions,

it may be better to continue its use while keeping in mind how it deviates from that which it is intended to capture. In this respect idealizations are like maps: They always require a compromise between overview and detail, and it is often advisable to use different maps for different purposes.

The reason why we idealize−perfect is that as philosophers we are at least as interested in what should be as in what is. Throughout the long history of our discipline, philosophers have tried to answer questions about how to think and how to behave. Requirements of rationality are usually important parts of the answers to such questions, and therefore idealization−perfection is commonly concerned with the ideal of rationality. Philosophical investigations of inferences, beliefs, decisions, and moral behaviour usually expound on the behaviour of rational thinkers, believers, decision makers and moral agents. We idealize−perfect in order to get a grip on what rationality demands of us, and sometimes also in order to gain insights on other normative demands such as those of morality.

It is important to keep track of one's idealizations and the reasons for them. Unfortunately, that is often not done. A particularly problematic confusion is that between the two forms of idealization. As one example of this, most accounts of human preferences depict them as transitive, i.e. someone who prefers a to b and b to c is assumed to also prefer a to c. That is not always the case for real-life preferences.[5] The reason why transitivity is assumed may be that the concept has been idealized−simplified, idealized−perfected, or both. In a discussion of divergences between the model and actual human behaviour it is important to know why the model assumes transitivity. Our analysis of such divergences may differ depending on whether transitivity was assumed for perfecting or simplifying reasons.

1.4.2 An Idealization in Two Steps

Formalization in philosophy typically results from an idealization in two steps, first from common language to a regimented philosophical language, and then from regimented into mathematical or logical language. For example, consider the derivation of the permission predicate (P) of deontic logic from the non-philosophical concept of a permission. We can use the following example from non-regimented language:

(1) Li-Hua is permitted to drive the forklift.

Here, the permission refers to an action. In regimented philosophical language, it is common to represent each action by a sentence denoting the state of affairs consisting in that action taking place. Hence:

[5] See Chap. 29.

(2) It is permitted$_{\text{phil}}$ that Li-Hua drives the forklift.

where "permitted$_{\text{phil}}$" is the philosophical idealization of the "permitted" of ordinary language. "Permitted$_{\text{phil}}$" differs from "permitted" in referring exclusively to what conscious agents do. It also differs in other ways. In non-philosophical usage, "when saying that an action is permitted we mean that one is at liberty to perform it, that one may either perform the action or refrain from performing it." In regimented philosophical language, however, "being permitted to perform an action is compatible with having to perform it" [67, p. 161].

The difference is perhaps best illustrated by the fact that in ordinary language we do not call something "permitted" that is in fact obligatory. Suppose that someone pays you in advance for cleaning their house. It would seem strange to say that you are then "permitted" to clean the house, since that would give the impression that you have a choice to do otherwise. However, according to philosophical usage of the term, it would be correct to say that you are permitted to do the cleaning. More generally, in philosophical language a permission is assumed to hold whenever the corresponding obligation holds ([67, p. 161]; [1, p. 55]; [10, p. 76]).

The second step of idealization takes us from "permitted$_{\text{phil}}$" to the deontic predicate P. This means that we go from (2) to

(3) Pa,

where P is a predicate expressing permission and a the sentence (or the proposition represented by the sentence) "Li-Hua drives the forklift". There are major differences in meaning between "permitted" and P. It should be noted, though, that in terms of most of the more philosophically significant differences, "permitted$_{\text{phil}}$" is closer to P than to "permitted". This applies for instance to the property of "permitted$_{\text{phil}}$" that we focused on above, namely that it holds for whatever is obligatory. This corresponds rather exactly the property of P that for all actions a, Oa implies Pa, where Oa is the corresponding predicate of obligation.

Intuitively speaking, most of the idealization in this example took place in the first step (from ordinary language to regimented philosophical language) rather than in the second (from regimented to formal language). And this is not untypical. Informal idealizations can sometimes be quite far-reaching. For instance, the concept of a person used in some philosophical discussions on personal identity is remarkably remote from the concept of a person in everyday language.

As all this should make clear, the difference between logical treatments of philosophical subject matter and treatments of the same matter in regimented natural language is *not* their distance to everyday concepts. The major differences are instead the mathematical skills that the formal models require and the characteristic types of questions that can be asked and answered with their help. Some philosophers who complain about the lacking realism of formal representations may to some extent confuse unfamiliarity in appearance with dissimilarity in meaning.

1.4.3 Selecting Concepts for Formal Representation

Logic is concerned with reasoning, but not all types of reasoning are included in the subject matter of logic. When discussing the pros and cons of different cars, we use arguments couched in terms such as "safe", "comfortable", "easy to drive", etc. These terms and their interrelations are not part of logic. Similarly, the terms used in wine tasting, such as "earthy" and "fruity" do not seem to have been subject to logical (or other formal) analysis. The same applies, of course, to the vast majority of terms that we use in different types of arguments. Logic is only concerned with a small fraction of the concepts and thought patterns employed in argumentation and reasoning. Whereas virtually every concept with some role in philosophy has been subject to some degree of informal idealization, only few of them have been formalized. Those that have been formalized are characterized by having wide usage and a role in inferences that is largely independent of context.

The core concepts of logic are the truth-functional concepts "and", "or", "not", "if. . . then", "if and only if", "some", and "all". These are concepts that we assume, for good reason, to have the same role in inferences in widely different contexts. They are often called the logical constants. (However, the definition of a logical constant is controversial [56].) But very few of the issues about valid argumentation that arise in philosophy (or outside of philosophy) concern the properties of words like these. It is more common for such issues to be concerned with the rules governing our usage of terms such as "know", "believe", "try", "do", "good", "better", "ought", "forbidden", and "permitted". These are also concepts with interesting structural interrelations that are fairly constant across contexts, and they have all been subject to logical formalization.

The choice of concepts for formalization should ultimately depend on whether the resulting models will be useful for philosophical and other worthwhile purposes. Since the shaping of new formal models is a creative rather than a rule-bound process, the following five critera for what to formalize should be read as tentative suggestions and nothing more.

First, the promising candidates usually have a meaning that is reasonably *constant across contexts*. This applies for instance to the words "good" and "bad". The meaning of these terms is presumably the same if we isolate them from a discussion on good and bad teachers as if we isolate them from a discussion on good and bad refrigerators [29]. This makes "good" and "bad" more promising candidates than, say, "earthy" or "sweet".

Secondly, the promising candidates usually provide a *structure* into which other, more context-specific, concepts can be inserted. This applies pre-eminently to the truth-functional concepts, but also to many others, such as our examples "good" and "bad". We can for, instance, talk about a collection containing both good and bad books, or an organ having both good and bad registers. Other examples are the action-theoretical concepts "do", "try", "refrain from", and "see to it that", to which we can affix more context-specific expressions denoting various types of actions.

Thirdly, it is usually a positive sign if we can combine the concept with some kind of *logical or other mathematical operations*. The most common examples are truth-functional and set-theoretical operations. As an example of the former, the concept "see to it that" can be combined with sentences describing states of affairs, and such sentences can be negated, combined into conjunctions and disjunctions, etc. As an example of the latter, in discussions about collective action we can talk about different groups of people, and one such group may for instance be a subset of another.

Fourthly, promising candidates tend to come with interesting issues about *potential structural properties* that seem to be generalizable across contexts. We can for instance ask whether something can be at the same time both good and bad. We can also ask whether someone who sees to it that a thereby also sees to that a-or-b.

Fifthly and finally, it is also a good sign if *connections with previous formalizations* are in sight. For instance, a logic of "good" and "bad" has obvious connections with the logic of "better". (If a is good and b bad, can we conclude that a is better than b?) Similarly, a logic of collective action can be connected with previously developed logics of individual action.

1.4.4 Structuralizing

After we have chosen a concept for formalization, we have to idealize it to make it suitable for formal treatment. As noted above, much of that idealization has often already been performed in informal philosophy. But for the purpose of formalization we may need to streamline the structural properties of the concept somewhat further. We can call this form of idealization *structuralizing*. In practice it often consists of the unification or splitting of concepts and the search for definability relations.

The *unification* of concepts is usually advisable when we are dealing with conceptually closely related terms in the informal language that have important structural properties in common. Such terms often differ in fine details that we cannot capture in the formal language without losing too much in simplicity. We have already seen one example that answers to this description, namely the unification of betterness and (converse) worseness. Another example is the collection of words used in ordinary language to denote obligatoriness: "must", "should", "ought", "have to", etc. These words are not exact synonyms. Typically, "obligations" originate from promises or agreements, whereas "duties" are associated with roles and offices in organizations and institutions [6, 18, 62]. Already in informal moral philosophy it is nevertheless common to regard "Yasmin ought to …", "It is a duty for Yasmin to …", and "Yasmin has an obligation to …" as synonymous. The reason for this is that the differences in meaning between these expressions have little or no relevance in most philosophical discussions. In deontic logic this simplification is even more useful. Therefore deontic logic standardly contains

only a single prescriptive predicate (denoted O) rather than several predicates corresponding to different prescriptive natural-language terms. The prescriptive predicate of the formal language can be seen as representing the common core of the various prescriptive expressions in natural language. Arguably, this core is more streamlined and more suitable for formal treatment than each of the natural-language predicates that were the starting-points of the formalization.

The opposite operation of *splitting* concepts is useful, sometimes necessary, when the concept we wish to formalize has meanings that differ in their structural properties. The splitting of concepts is, of course, quite common also in informal philosophy, but in preparing for formalization we have to pay particular attention to structural properties when deciding whether or not to split a concept.

Again, we can use prescriptive terms from moral philosophy as examples. Consider the following two sentences:

(a) "You must help her."
(b) "You must be wrong."

(a) expresses an obligation. (b) does not. Instead it expresses necessity. This is reason enough for the informal philosopher to distinguish between the two meanings of the word. For the formal philosopher there is an additional reason, namely that the two senses have different structural (logical) properties. To see this, consider the following property:

If $Must(X)$ then X.

This property holds for the "must" of our second example. If I am right in saying that you must be wrong, then surely you are wrong. We can easily verify that the property also holds in other cases where "must" is used in the same sense. But it does not hold in the first example. Even if I am right in saying that you must help the person referred to, it certainly does not follow that you actually do so. Again, we can verify that the same applies to other sentences where "must" has the same meaning. Such a consistent difference in terms of (logical) structure is a sure sign that for the purposes of formalization, "must" has to be split into two concepts. It is only obligation−must that can be unified with the other prescriptive predicates into the deontic operator O. Necessity−must can instead be unified with "necessary", "unavoidable" and the like.

Next, consider the following two uses of the word "ought":

(c) "You ought to help your destitute brother."
(d) "There ought to be no suffering in the world."

(c) expresses a prescription, something that someone should do. Alternatively, we could express the same statement with some other such term, saying for instance "You have a duty to help your destitute brother". In this respect, (d) is quite different. It expresses a wish about the state of the world, or an evaluation of such a state. It does not directly prescribe or recommend any action. This is a well-known distinction. The "ought" of (c) is called ought-to-do (Tunsollen) and that of (d) ought-to-be (Seinsollen or ideal ought) ([68, p. 195]; [13]).

This double usage is specific for "ought", and does not apply to prescriptive predicates in general. It would not make much sense to say that there is a duty for the world not to contain any suffering. Since deontic logic is concerned with prescriptions in general, not only those expressed by the English word "ought", ought-to-be and ought-to-do have to be split. Only the latter should be unified with the other prescriptive predicates into the deontic O operator. Just like necessity—must, ought-to-be should be treated as a separate concept, not to be merged or confused with the prescriptive ones.

Unfortunately, this has not always been realized. (Perhaps one of the reasons for this is that the O operator is usually read "ought", and we are not sufficiently often reminded that in spite of this, it represents the common core of several natural language expressions.) A considerable amount of confusion has been created by attempts to unify ought-to-do with ought-to-be. This is usually done by reconstructing ought-to-do as ought-to-be referring to actions, in the way shown in the following two examples:

Person i ought to do x. = It ought to be the case that person i does x.
Person i ought to do x. = The world ought to be such that person i does x.

But this does not work. "You ought to sing in tune" means something quite different from "The world ought to be such that you sing in tune." And more generally speaking, that which we ought to do does not coincide with that which the world ought to be such that we do. The world ought to be free of racism, and in such a world no one would help victims of racism (since there would be none). Recently, a newly wed woman was killed by a robber. It certainly ought not to be the case that her husband went to her funeral less than a month after they married. But of course he ought to go to the funeral. The distinction between ought-to-do and ought-to-be is fundamental, and the two notions should be kept apart in both formal or informal moral discourse [35].

If one concept is *definable* in terms of another, then we can focus on the latter, and treat the defined concept as a mere abbreviation in the formal language. It is not uncommon for philosophically important concepts to be definable in terms of each other. One example is the interdefinability among the three modal concepts of necessity, possibility and impossibility. To be impossible means not to be possible, and something is necessary if and only if it is not possible that it is not the case. Letting $\Box$ stand for necessity, $\Diamond$ for possibility, and $\nDiamond$ for impossibility, we can express these relationships as follows:

$$\Box a \leftrightarrow \nDiamond \neg a$$

$$\Box a \leftrightarrow \neg \Diamond \neg a$$

$$\nDiamond a \leftrightarrow \neg \Diamond a$$

It follows, of course, that in a modal logic we can take any of these three concepts as primitive (undefined), and define the others in terms of it. It is unimportant which of them we select to be the primitive notion.

In other cases, definability comes only in one direction. We can for instance define "best" in terms of "better" in the following way:

x is (uniquely) best if and only if for all y other than x: x is better than y.

However, there is no corresponding way to define "better" directly in terms of "best".[6] Therefore, "best" is in practice always treated as a defined concept in formal languages.

In general, logical languages with fewer primitive (undefined) concepts tend to be more manageable. The aim to have as few primitives as possible is called *definitional economy*. In order to achieve it we have to investigate carefully if some of the concepts on our agenda for formalization can be defined in terms of some of the others.

1.4.5 Introducing Formulas

As we have already seen, the concepts that are subject to formalization tend to owe much of their usefulness to the ways in which they can be connected to various more specific expressions. The common truth-functional connectives can be combined with any sentences carrying truth-values. To the deontic operators P and O we attach action-describing sentences. To a "stit" (see-to-it-that) operator we connect a name representing a person and a sentence describing a potential outcome of an action by that person, for instance:

$stit_i a$

where i is a person and a the outcome of that person's action. These attachments are called "variables".

Variables are essential components of formal languages; without them non-trivial formalization would not be possible. Historically, they are an important invention. In medieval times, names (such as "Socrates") were used to denote arbitrary persons. That practice is still frequent in philosophical texts, but it is also common to use single letters to denote persons. ("If A borrows money from B and then gives it to C,. . . .") Informal philosophical discourse also contains symbols representing objects that do not have proper names in other contexts. ("If the state of affairs a obtains at time t,. . . .") In logic, we do more of the same. The following series of synonymous statements illustrates the different degrees of compactness of notation:

[6] We can do so if we manipulate the sets of alternatives, see Chap. 27.

Ordinary language:
The first cause took place either before or at the same time as the second cause, and the second cause took place before the effect.

Informal philosophical language:
c_1 preceded or was simultaneous with c_2, and c_2 preceded e.

Logical language:
$c_1 \leq c_2 \ \& \ c_2 < e$.

The following series of restatements of a definition illustrates the pros and cons of the more compact notation that formalization makes available.

Ordinary language:
A cousin is a person with whom one has at least one grandparent in common but no parent in common.

Semi-formal language 1:
Person i is a cousin of person j if and only if (1) there is a person who is a grandparent of both i and j, but (2) there is no person who is a parent of both i and j.

Semi-formal language 2:
Person i is a cousin of person j if and only if (1) there is a person x who is a grandparent of both i and j, but (2) there is no person v who is a parent of both i and j.

Semi-formal language 3:
Person i is a cousin of person j if and only if (1) there are persons x, y, and z such that x is a parent of y who is a parent of i and x is also a parent of z who is a parent of j, but (2) there is no person v who is a parent of both i and j.

Logical language:
iCj if and only if :
$(\exists x)(\exists y)(\exists z)(xPyPi \ \& \ xPzPj) \ \& \ \neg(\exists v)(vPi \ \& \ vPj)$

The first of these statements is clearly the most easily readable one, and the last is the most precise and compact one. For many purposes, some compromise between readability and precision may be desirable; then one of the intermediate, semi-formal options may be optimal. The cases when logical notation is most useful are those in which we want to prove some property of the concepts we are working with. Box 1.1 on page 25 shows how the compactness of formal notation makes a proof easier to follow.

Box 1.1 Two versions of the same argument
Consider the two relations on points in time: "precedes or is equal to" ($\leq$) and "precedes" ($<$). We are going to show that if the former of these is transitive, then so is the latter.

In natural language

Consider three points in time such that the first precedes the second and the second precedes the third. Then clearly the first precedes or is equal to the second, and the second precedes or is equal to the third. Since the relation "precedes or is equal to" is transitive, we can conclude that the first precedes or is equal to the third. Now suppose that the first does not precede the third. Since the first precedes or is equal to third, we can conclude that the third is equal to the first. Thus the third precedes or is equal to the second. But this is impossible since the second precedes the third. We have derived a contradiction from the assumption that the first does not precede the third. Thus the first precedes the third. This shows that the relation "precedes" is transitive.

In formal language

Let t_1, t_2, and t_3 be three points in time such that $t_1 < t_2$ and $t_2 < t_3$. Then $t_1 \leq t_2$ and $t_2 \leq t_3$, and transitivity yields $t_1 \leq t_3$. Now suppose that $t_1 < t_3$ is not the case. It then follows from $t_1 \leq t_3$ that $t_1 = t_3$. We can then substitute t_3 for t_1 in $t_1 \leq t_2$, and obtain $t_3 \leq t_2$. But that is impossible since $t_2 < t_3$. It follows from this contradiction that $t_1 < t_3$. [36]

1.4.6 Determining the Number of Variables

In ordinary language, one and the same concept can be associated with different numbers of variables:

Cynthia is a mother.
Cynthia is Peter's mother.

It would be tempting to follow the same pattern in formal language, and (with the predicate M denoting motherhood) translate the sentences as follows:

Mc
Mcp

This would require that we allow one and the same predicate to appear with different numbers of variables. However, the introduction of such flexible predicates would

Table 1.1 Four usages of the term "free" that differ with respect to the variables

Example	Schema
She is free now.	i is free.
She is free from all those debts now.	i is free from the obstacle x.
Finally she was free to take up her studies again.	i is free to perform the action y.
She is now free from any legal obstacles to leave the country.	i is free from the obstacle x to perform the action y.

leave the exact relationship between formulas such as Mc and Mcp unclear. There is a much better way to deal with this, namely to introduce M as a two-place predicate, which means that an expression containing M can only be well-defined if each instance of M has two variables. The single-variable expression "Cynthia is a mother" is synonymous with "Cynthia is someone's mother", which we can express with the existential quantifier ∃ as follows:

$$(\exists x)Mcx$$

When introducing a predicate or a relation into the formal language, it is important to choose the right number and type(s) of variables. It is often preferable to include representations of all the variables that can be attached to the corresponding informal expressions, and then define uses with a reduced number of variables in the way we just did for motherhood.

The term "free" as used in political philosophy is an interesting example of this. If we classify uses of "free" in informal language according to the variables, then we will find at least four variants. Table 1.1 gives examples of these, and it also provides general schemata for each of the variants. These variants represent different notions of freedom, notions that are controversial in political philosophy. Some political thinkers have claimed that all true freedoms can be fully expressed by statements of the second type, "freedom from" (negative freedom). Others have put much emphasis on freedoms representable by the third type of expressions, "freedom to" (positive freedom). They see freedom largely as ability to make and implement one's own choices [4]. The fourth variant is less common, but it is quite useful since all the others can be defined in terms of it [55]. In formal analysis it would take the form of a three-place predicate

$$F(i, x, y)$$

where i is an individual, x an obstacle, and y some action that the individual can potentially perform. In this case it is much more difficult than for motherhood to determine how the three-place predicate should be used to define the two-place and one-place ones. As a first attempt we could define "freedom to" as $(\forall x)F(i, x, y)$, i.e. one is free to y if and only if one is free from all obstacles that might prevent the attainment of y. However, that may seem somewhat extreme. Arguably I am free

Table 1.2 Two usages of the term "duty" that differ in terms of the variables

Example	Schema
It is his duty to answer the phone on all times of the day.	Person i has a duty to do x.
The lawyer has a duty towards the client to defend her interests.	Person i has a duty towards person j to do x.

to read the morning newspaper even if a snowdrift makes it impossible for me to get hold of it. A distinction between different classes of obstacles may have to be introduced. Similar problems arise for the reduction of the three-place predicate to the two-place "freedom from". However, these difficulties should not be counted against the three-place predicate. To the contrary, these are real philosophical difficulties in the analysis of political freedom. The three-place predicate is a tool to present these difficulties more clearly, thereby making them more amenable to precise analysis.

But this is a controversial area. Traditionally, the negative notion of freedom is associated with right-leaning and the positive notion with left-oriented political ideas. Not surprisingly, the three-place predicate has been accused of both a left-wing and a right-wing bias ([22]; [64, p. 253]). Nevertheless, it has the advantage of allowing us to represent "freedom from" and "freedom to" in one and the same format, rather than just treating them as mutually incompatible notions.

In doubtful cases it is usually better to include than to exclude a variable when introducing a formal predicate. But of course, there are cases when one or other of the variables has such a small role that it can for most purposes be excluded. For a possible example, let us consider the notion of a duty, as shown in the examples and schemas of Table 1.2. Common usage of the term "duty" is dominated by the first variant mentioned in the table, two-place duty. The second variant, the three-place notion of a duty, is more uncommon. The two-place notion has the advantage that it can be unified with other prescriptive notions in the way discussed above. (Some of these, such as "morally required", do not have a three-place variant.) It is indeed common practice in philosophy to treat duty as a two-place concept. There are good reasons for this practice, but it has a price: We lose the ability to express that someone owes something to a specific person. Such relationships will then have to be treated in separate investigations, using a different formal representation [30, 58]. As noted above, there is nothing wrong with using different formal representations of a concept for different purposes.

But something more can be learned from this example. Even the two-place format "Person i has a duty to do x" does not correspond to the standard deontic operator for obligations, namely Ox which only has place for one variable. How is that possible? Obligations are normally tied to persons, and surely it makes a difference who is subject to an obligation? The explanation is that x in Ox is normally taken to refer to an action by a specified agent. If x represents the action consisting in me paying my rent, then we can take it for granted that I am the duty-holder in Ox. However, this is a rather precarious principle since information that is

not stated explicitly runs a risk of being forgotten or misunderstood. The suppression of the person variable can make us forget about its existence, so that we treat moral prescriptions as impersonal although they are not. This may be one of the sources of the confusion about ought-to-be that was referred to above.

1.4.7 Specifying the Domains of the Variables

For each variable-place attached to a predicate we need a well-defined domain (source), i.e. a set whose elements represent the objects that variables in that place can stand for. In some cases the same domain can be used for more than one variable-place. This applies to the two-place predicate of motherhood. Here we can use the same domain, namely the set of all human beings, for both variable-places. For the three-place predicate of freedom the situation is quite different. We need three sets of variables, representing persons, obstacles, and actions.

In a formal treatment it is important to assign well-defined domains to all variable-places, and to be careful not to transgress them. There are two ways to deal with the complication that different variable-places refer to different groups of objects. To exemplify this, consider a simple logic of parenthood relationships with the predicates F and M, such that Fxy means that x is father of y and Mxy that x is mother of y. We can assume that fathers are men, mothers are women, and their children can be either. One way to express this is to use two sets of variables, $\mathbb{W}$ representing women and $\mathbb{M}$ representing men, and then introduce the distinction in the requirements for formulas to be well-formed, as follows:

Mxy is a well-formed formula if and only if $x \in \mathbb{W}$ and $y \in \mathbb{M} \cup \mathbb{W}$.
Fxy is a well-formed formula if and only if $x \in \mathbb{M}$ and $y \in \mathbb{M} \cup \mathbb{W}$.

The other alternative is to have only one domain, namely the domain $\mathbb{H}$ consisting of human beings, and include the restrictions in the logic rather than in the formation rules for the language. This can be done with one-place predicates denoting "is male" and "is female":

Each of Mxy and Fxy is a well-formed formula if and only if $x \in \mathbb{H}$ and $y \in \mathbb{H}$.
From Mxy it follows logically that Lx, where L denotes "is female".
From Fxy it follows logically that Gx, where G denotes "is male".

The two approaches are equivalent, and the choice between them is a matter of taste and convenience. The latter approach places the restrictions in the logic rather than in the language. This can be seen as an advantage since it makes the restrictions somewhat more accessible to modifications and adjustments. For instance, if $(\forall x)(Lx \vee Gx)$ holds in our original statement of the logic, then we can easily remove this principle in order to include people who are neither female or

male. It may be an advantage to be able to do this without changing the language, which is considered to be a more drastic change of the framework.

1.5 Building a Logical Language

The distinguishing feature of logical language as compared to other formal languages is its focus on the representation of propositions (statements), by which we usually mean something that can be either true or false. In natural language we express propositions with sentences. One and the same proposition can be expressed by different sentences. Thus, "Dana is married to Lou" expresses the same proposition as "Lou is married to Dana" (and of course the same proposition can also be expressed by sentences in other natural languages).

Other formal languages than logic also contain sentences expressing propositions. In the appropriate contexts, $x^2 = y^2 + z^2$ represents a proposition about the relationships between the lengths of the hypotenuse and the legs of a right-angled triangle, $E = mc^2$ one about the mass−energy equivalence in relativity theory, etc. However, logic is distinguished by the generality of its treatment of sentences and by its suitability for formal work related to the conclusions that can be drawn from sets of sentences.

Sometimes, logical expressions are used to represent sentences that are not to be classified as true or false, but rather according to some other dichotomy, such as that between morally approved and morally unapproved actions or states of affairs. There are also logical systems, called many-valued logics, in which the traditional true/false dichotomy is replaced by a classification containing more than two alternatives, such as true/false/unknown. These distinctions have little impact on the construction of logical languages, and they will therefore not be considered in this section.

The simplest tools for building a logical language are those that treat sentences as wholes and do not contain separate representations of their parts. These constructions will be the topic of Sect. 1.5.1. In Sect. 1.5.2 we turn to the construction of sentences from their parts, and in Sect. 1.5.3 to formal elements that refer to the parts of the sentences thus formed. Section 1.5.4 shows how the formation rules for a formal language are usually expressed.

1.5.1 From Atomic to Composite Sentences

Let us start with a set of (proposition-representing) sentences. We can call them $a, b, \ldots$. To begin with, we will treat them as "atoms" ("atomic sentences"), i.e. we disregard their internal structure. This is of course a choice of a level of abstraction.

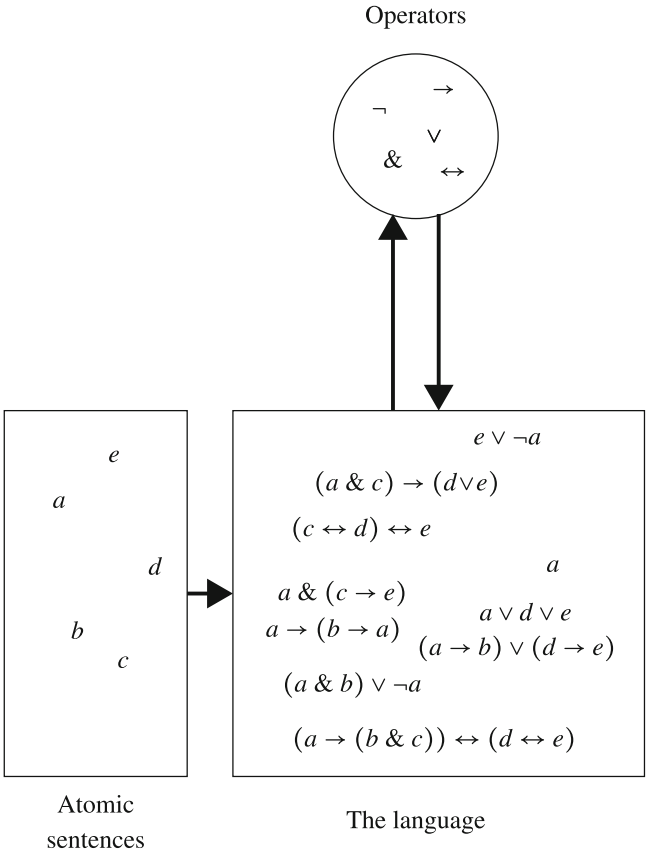

Fig. 1.1 A language formation diagram for sentential (propositional) logic

It provides us with a sort of bird's-eye view that has turned out to be quite useful for the study of phenomena such as conclusions and assumptions.

In order to get things going we need means to combine atomic sentences to form composite, or as we usually say, molecular sentences. The construction elements used for this purpose are called sentential operators, since they are operators that take us from a sentence (or several sentences) to a new sentence. The simplest sentential operator is negation, often denoted ¬. It takes us from a sentence a to its negation $\neg a$. If a represents the same proposition as "I am tired", then $\neg a$ represents the same proposition as "I am not tired". Other such operators are conjunction ("and", &), disjunction ("or", $\vee$), material implication ("if ... then", $\rightarrow$ or $\supset$), and equivalence ("if and only if", $\leftrightarrow$ or $\equiv$). (All these are truth-functional operators, but that is not a property of the language but one of the logic.)

Figure 1.1 shows how these operators can be used to form the full language of propositional logic. Two important features should be noted in that diagram. First, the atomic sentences are themselves directly introduced into the language, as shown

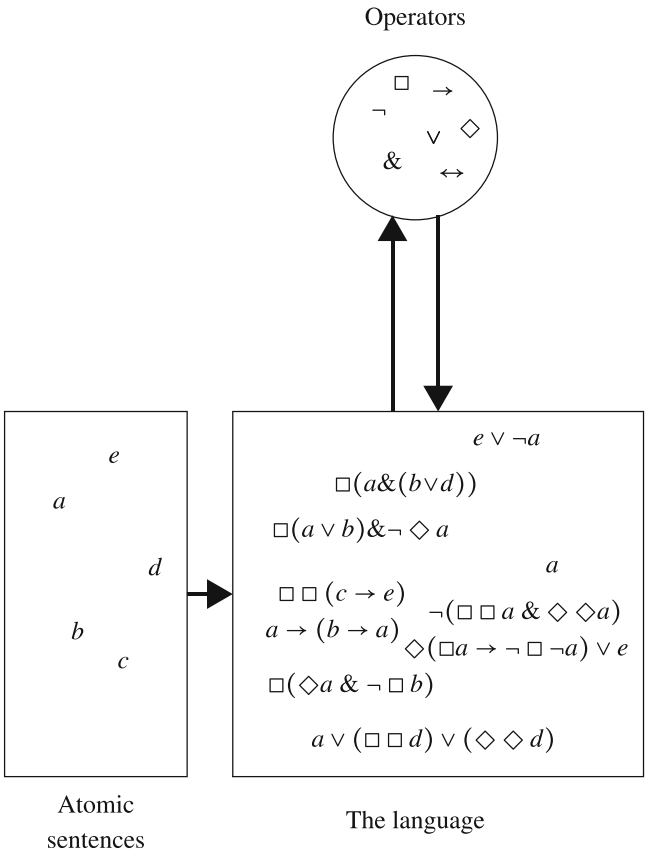

Fig. 1.2 A language formation diagram for modal logic. □ stands for necessity and ◊ for possibility

with the horizontal arrow. Secondly, the operators ¬, & etc. can be applied not only to atomic but also to molecular formulas. This means that unlimitedly complex formulas can be formed, such as $\neg a \vee \neg(b \vee c)$, etc.

Other operators can be added to the language in the same way. In a discussion about necessity and possibility we will need the unary (single input) operators □ ("it is necessary that ...") and ◊ ("it is possible that ..."), and often also the binary (two input) operator of strict implication ⇒ ("if ... then necessarily ..."). These are inserted into the logical language in Fig. 1.2. An important feature of this language is that □ and ◊ can take as inputs sentences in which they are themselves already present. We can therefore form sentences such as $\Box\Box b$ and $\Box(a \rightarrow \Diamond(a \vee b))$. From an interpretational point of view this is not quite uncontroversial. It can for instance be questioned whether a sentence such as $\Box\Box b$ ("it is necessary that it is necessary that b") is at all meaningful. Is necessity iterable?

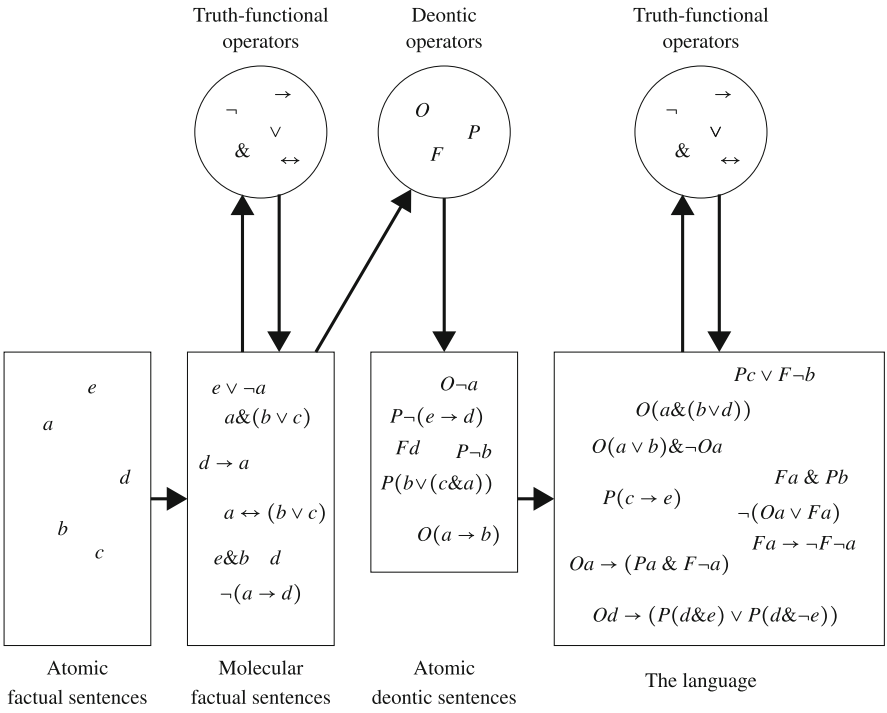

Fig. 1.3 A language formation diagram for a deontic logic that does not allow the iteration of deontic operators. O stands for obligation, P for permission, and F for prohibition

One operator whose repeated use in a formula has often been questioned is the deontic operator O that stands for moral requirement. From a factual statement a representing some human action we can form the sentence Oa saying that a is morally required. But how meaningful is the sentence OOa? Does it say that it is morally required that it is morally required that a? Then, exactly what does that mean? There are reasonable interpretations of O that make this sentence meaningless. In order to block the formation of such sentences we need to construct a somewhat more complex language formation diagram, as shown in Fig. 1.3. Here we are not allowed to affix O to sentences already containing O. Therefore neither OOa nor $O(Oa \vee O\neg a)$ are well-formed formulas, which means that although they consist of parts of the language, they are not themselves parts of the language. However, we can apply truth-functional operators to sentences containing O, forming sentences such as $\neg O(a \to b)$ and $Oa \vee Ob$.

Figure 1.4 shows an alternative language formation diagram for a deontic language that does not allow "repeated" application of the deontic operators. The difference between Figs. 1.3 and 1.4 is that in the latter, atomic and molecular factual statements such as a, $a \vee b$ etc. are directly included in the deontic language. Here

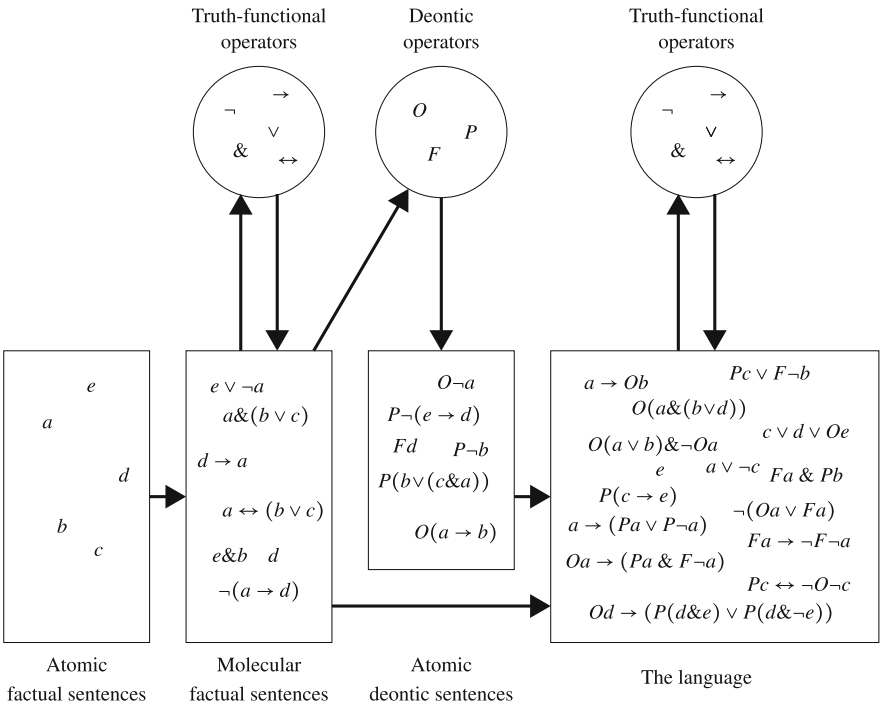

Fig. 1.4 Another language formation diagram for a deontic logic that does not allow the iteration of deontic operators. It differs from the diagram of Fig. 1.3 in allowing for the direct introduction of factual sentences into the language

it is also possible to form sentences such as $Op \ \& \ \neg p$ and other sentences with "mixed" deontic and factual contents.

1.5.2 Decomposing the Atoms

Factual sentences in many natural languages tend to have a standard grammatical form containing two main parts, a subject and a predicate. The subject represents that which we say something about, and the predicate that which we say about it:

Socrates	wrote no book.	Nevertheless,	his thoughts	changed the world.
subject	*predicate*	*connective*	*subject*	*predicate*
sentence 1			*sentence 2*	

As we have already seen, formal logic has taken over this structure from natural language. Predicate language that is based on the subject/predicate distinction is by far the most common formal representation used to decompose the logical atoms and scrutinize their components. When we translate the sentence "The author is

a bore" into predicate logic, we identify the subject ("The author") and assign a symbol such as i to it. Similarly, we identify the predicate ("is a bore") and assign to it a symbol such as B. The sentence is then denoted Bi.

Many natural language predicates refer to some specific individual or other object of thought. In grammars of natural language there are two competing ways to analyze such sentences:

First analysis:		*Second analysis:*		
Angelina	is in love with Barbara.	Angelina	is in love with	Barbara.
subject	*predicate*	*subject*	*predicate*	*object*
sentence		*sentence*		

For the purposes of logic the second analysis is preferred, since it allows for more detailed investigations. It makes use of the predicate "is in love with" which takes two variables, one of which corresponds to the subject and the other to the object of the natural language sentence. The sentence can then be rendered by the formula Lab, where L represents "is in love with".

As mentioned in Sect. 1.4.6, each predicate always takes the same number of variables (often called arguments). The number of variables is often called the arity of the predicate. A predicate is called unary (monadic, 1-ary, one-place) if it takes one variable, binary (dyadic, 2-ary, two-place) if it takes two, ternary (3-ary, three-place) if it takes three, and for any natural number n it is called n-ary (n-place) if it takes n variables. A 0-ary (nullary, zero-place) predicate, i.e. a predicate without variables, functions in the same way as an atomic sentence. That a predicate is nullary does not mean that there is nothing that it says something about. Instead, this means that we have chosen not to decompose it and introduce variables representing one or more of its components.

When formalizing natural language, it is a good general rule to use predicates with the lowest arity that is compatible with an adequate representation of the subject matter. In particular, if predicates with high arity can be replaced by truth-functional combinations of predicates with lower arity, then that should be done. "The author and the bookseller are bores" should be translated as $Ba \,\&\, Bb$, not as $\overline{B}ab$ with a dyadic predicate $\overline{B}$. Similarly, "Ivan and Joanna are Kelly's parents" should be translated as $Pik \,\&\, Pjk$, not as $\overline{P}ijk$.

Two warnings are warranted. First, use nothing else than the existential quantifier ($\exists$) to represent "exists". Do *not* introduce a predicate to represent "exists", since doing so gives rise to complications that you would like to avoid [8]. Secondly, use nothing else than the equality sign ($=$) to represent "is equal to" or "is the same as" as a binary predicate. It is important to follow the standard rules for predicate logic with identity, which can be found in most textbooks on elementary logic.

Figure 1.5 shows a formation diagram for a simple predicate language with monadic and dyadic predicates. Note that the components of sentences (variables and non-nullary predicates) are not themselves included in the language, contrary to the atomic sentences in Figs. 1.1, 1.2, and 1.4.

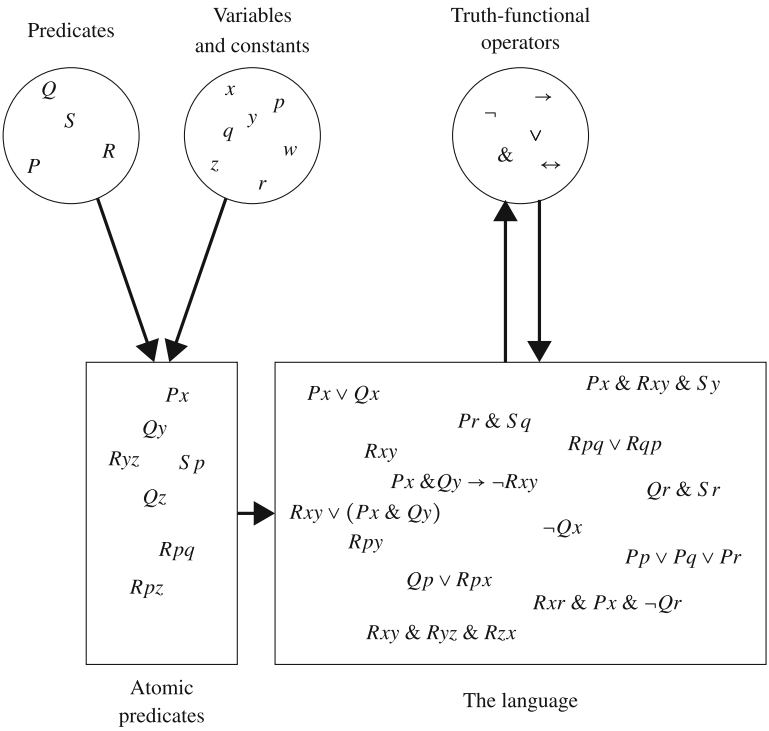

Fig. 1.5 A language formation diagram for standard predicate logic without quantifiers

1.5.3 Quantifiers

In order to make efficient use of the decomposition of atomic sentences into predicates and variables, we need to employ Frege's great invention, quantifiers. The quantifiers ∀ ("all") and ∃ ("some") are a type of sentential operators. Just like the monadic operators referred to in Sect. 1.5.1, they take us from a sentence to another sentence, hence if Fxy denotes "x has y as a friend" and i denotes the author, then $(\exists y)Fiy$ says that the author has some friend. Similarly, the sentence

$$(\forall x)(\forall y)(Fxy \rightarrow Fyx)$$

says that friendship is always mutual, whereas the sentence

$$(\exists x)(\exists y)(Lxy \,\&\, \neg Lyx)$$

where L denotes "loves" expresses the most unfortunate fact that the same does not apply to love.

The translation of sentences from natural language into predicate logic is not always straightforward, and sometimes it requires considerable changes in structure. Often, sentences with one and the same structure in natural language require quite different translations:

The dog is a Mastiff.
The giraffe is a mammal.

The first sentence is preferably translated into

$$Md$$

where M is the predicate "is a Mastiff" and d the particular dog referred to. The second sentence is best translated as

$$(\forall x)(Gx \rightarrow \overline{M}x)$$

where G is the predicate "is a giraffe" and $\overline{M}$ the predicate "is a mammal".

In a language with quantifiers we need to distinguish between constants and variables. A constant, such as d in our formula Md, refers invariably to a particular object, and it is not affected by quantifiers. It can be compared to a unique name such as "Louis XIV" in natural language. A variable, such as x in our formula $(\forall x)(Gx \rightarrow \overline{M}x)$, has no meaning in itself but acquires meaning in the context, just like pronouns such as "that", "this", and "it" in natural language.

The use of variables makes predicate logic well suited to keep track of complex relationships. The resources of natural language are much less suited for that purpose. We can distinguish between "this" and "that", but we do not use them repeatedly with persistent reference. We can introduce phrases like "the first person" and "the second person", but talk using such expressions is usually difficult to follow. (See Box 1.1 on p. 25.)

The language formation diagram in Fig. 1.6 (an extension of Fig. 1.5) summarizes the construction of predicate logic with quantifiers. Note that in this language, quantifiers cannot be applied to predicates. For instance, a formula such as $(\exists P)(\forall x)Px$ is not well-formed. Due to this limitation, the logic based on this language is called *first-order* predicate logic. In second-order predicate logic, $(\exists P)(\forall x)Px$ is a well-formed formula. (It can be interpreted as "There is a property that everything has.")

Ordinary language contains many expressions that have similar functions in sentences as "all" and "some":

Most Icelanders understand Norwegian.
Very few Germans understand Chinese.
At most three Government members have experience of blue-collar work.
There are *infinitely many* prime numbers.
The committee has an *an odd number of* members.

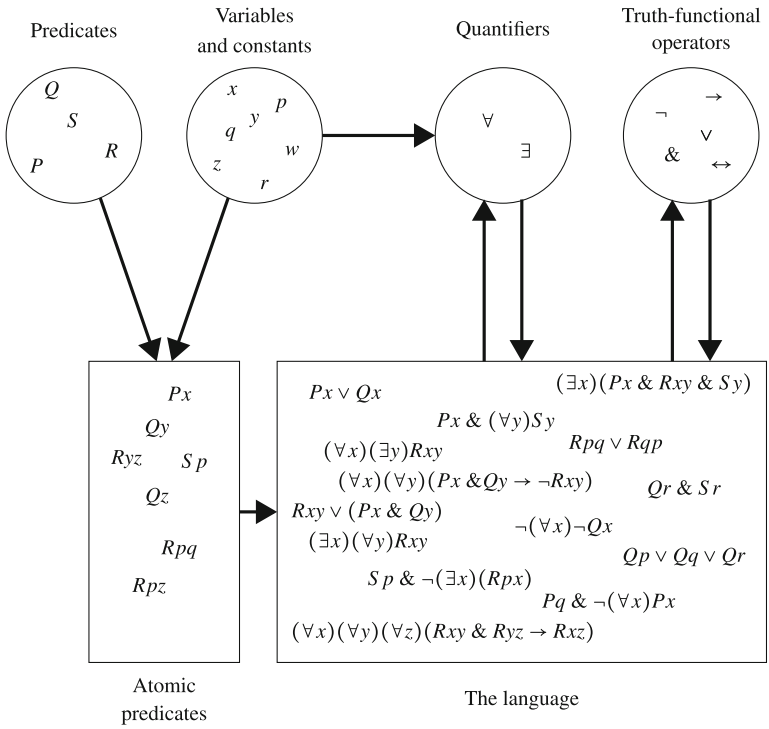

Fig. 1.6 A language formation diagram for standard predicate logic with quantifiers

The formal representation of such *generalized quantifiers* is of considerable philosophical interest. The same applies to second- and higher-order logics, in which the predicates themselves are treated as variables of quantifiers.

1.5.4 Specifying the Language

Formal languages are usually defined recursively, i.e. the definition identifies their smallest elements and then proceeds to specify how these elements can gradually be combined into larger and larger linguistic expressions. Our language formation diagrams show how this recursive process proceeds with repeatable steps of concatenation. In the specialized literature, this is expressed in more compact fashion. There are two common ways to specify a logical language. One is to list a set of language formation rules. In the following example this is done for the modal language presented in Fig. 1.2:

$\mathcal{L}$ is the language consisting exactly of the sentences obtainable through the following rules:

1. $T \subseteq \mathcal{L}$, where $T = \{a_1, a_2, \ldots\}$ is a countably infinite set of sentences.
2. If $\alpha \in \mathcal{L}$, then $\neg\alpha \in \mathcal{L}$, $\Box\alpha \in \mathcal{L}$, and $\Diamond\alpha \in \mathcal{L}$.
3. If $\alpha \in \mathcal{L}$ and $\beta \in \mathcal{L}$, then $\alpha \,\&\, \beta \in \mathcal{L}$, $\alpha \vee \beta \in \mathcal{L}$, $\alpha \rightarrow \beta \in \mathcal{L}$, and $\alpha \leftrightarrow \beta \in \mathcal{L}$.

The other method, particularly common in computer science, is an abbreviation of the former as a so-called Backus-Naur Grammar clause:

$$\phi ::= a_1, a_2, \ldots \mid \neg\phi \mid \phi \,\&\, \psi \mid \phi \vee \psi \mid \phi \rightarrow \psi \mid \phi \leftrightarrow \psi \mid \Box\phi \mid \Diamond\phi$$

Here, ::= denotes that the symbol to the left should be replaced by one of those on the right, and | denotes a choice among different such substitutions.

1.6 The Uses of Logical Inference

Translations into logical language can to some extent be clarifying in themselves. This applies for instance to translations from the rather erratic quantifiers of natural language to the more regular ones of predicate logic. But the most important advantages of formalization are only obtainable when we go beyond mere translation, and investigate, with logical tools, the properties of the models that we have built. It is a major advantage of formal models that they are so precisely described that such properties can be determined with certainty. Their major disadvantage, of course, is that these properties may be different from those of that which they are a model of. Efficient use of formal models requires both that we investigate the formal properties of our models *and* that we critically evaluate how these properties relate to those of the phenomena that led us to develop the models.

This is not the place to delve into the methodology of logico-mathematical work, how to construct axioms and prove lemmas and theorems. Instead, this section is devoted to the connections between the construction of a system of logical inferences and the process of formalization. Section 1.6.1 discusses the choice between extensional and non-extensional logic for sentential operators. Section 1.6.2 shows how logical analysis can reveal distinctions that are less obvious in natural language, thereby contributing to the development of new philosophical concepts. This is followed by a discussion of how logical analysis can lead to improvements of the formal framework itself. Sometimes minor adjustments are sufficient (Sect. 1.6.3). On other occasions, logical analysis forces us back to the drawing-board in search for a better formal model (Sect. 1.6.4).

Box 1.2 Two ways to use a name
"Charles-Édouard Jeanneret" was the legal full name and "Le Corbusier" the pseudonym of a famous architect. Consider the following sentences:

(1) Le Corbusier was born in 1887.
(2) Charles-Édouard Jeanneret was born in 1887.
(3) Le Corbusier is a pseudonym.
(4) Charles-Édouard Jeanneret is a pseudonym.

In the first two sentences, the names refer to the person. These two sentences have the same truth conditions, and they are indeed both true. In the last two sentences, the names refer to themselves, and these two sentences do not have the same truth conditions. (3) is true and (4) false.

In sentence (1), "Le Corbusier" is used *extensionally*, by which is meant that the truth-value of this sentence is not changed if "Le Corbusier" is replaced by another expression with the same extension. (The extension of an expression is the collection of objects to which it refers, in this case a collection consisting of one person.) In sentence (3), "Le Corbusier" is used non-extensionally.

1.6.1 Intersubstitutivity of Logical Equivalents

It is important in philosophy to distinguish between extensional and non-extensional uses of an expression. (See Box 1.2 for a reminder.) Therefore, when constructing the logic of a sentential operator, we have to decide whether to give it an extensional or a non-extensional logic.[7] We can illustrate this with the sentences about the Dodo (*Raphus cucullatus*) that can be formed with the following notation:

d The Dodo is extinct.
r *Raphus cucullatus* is extinct.
E There is sufficient scientific evidence that ....
K Alix knows that ....

Since r and d are equivalent, so are Er and Ed. More generally speaking, if we can replace a sentence attached to the operator E by an equivalent sentence, then the truth-value is not changed. An operator with this property is said to be *extensional* or satisfy *intersubstitutivity of logical equivalents*.

The operator K does not have this property, since Kr and Kd are not logically equivalent. It is both possible and quite common to know that the Dodo is extinct

[7] Rudolf Carnap [9, pp. 57–63] claimed that all non-extensional concepts can be reconstructed as extensional, but his mode of reconstruction has not caught on and does not seem to be practicable.

without knowing that *Raphus cucullatus* is extinct. This is a feature that K shares with most other operators representing attitudes such as believing, doubting, wishing, preferring etc.

When constructing a logical system that contains sentential operators, it is important to specify which of these operators satisfy intersubstitutivity and which do not. (Note that this is a property of the inference pattern applied to the language, not a property of the language itself.) It might seem obvious that a logical operator that represents a non-extensional concept in natural language, such as a knowledge or belief operator, should also be non-extensional. In practice, however, it is quite common to use extensional operators to represent non-extensional concepts. The reason for this is that non-extensionality usually comes with a high price: it makes the logic of the operator so weak that very little can be proved. Intersubstitutivity of logical equivalents is an idealization that allows us to have a much richer logic to work with. There are two major ways to justify that idealization.

The most common justification is that the "non-extensional" uses of the concept can in most cases easily be identified. We can therefore use an operator with an extensional logic and just bear in mind that it is inadequate to deal with problems where non-extensional properties of the underlying concept have a role. We can for instance develop a logic of belief with an extensional belief operator B and assign properties to it such as $Ba \,\&\, Bb \rightarrow B(a\&b)$ and $Ba \rightarrow \neg B\neg a$. A disadvantage with this approach is that the outer limits of the logic's area of application cannot be specified in precise terms.

The other, somewhat more sophisticated, approach is to change the interpretation of the operator so that it does not refer to the original ordinary-language concept but to some variant of it that can be expected to allow for the substitution of logical equivalents. For the belief operator such a reinterpretation has been proposed by Isaac Levi [50, 51]. His solution is to interpret B as referring to what the agent is committed to believe rather than what she actually believes.[8] A person who believes in the above statement d ("The Dodo is extinct"), is also committed to believe in r ("*Raphus cucullatus* is extinct") upon understanding its meaning. This approach has the advantage over the previous one that the delimitation is more precise and therefore more accessible to criticism and improvement.

For another example, consider again the predicate O of moral requirement. In deontic logic, O is usually taken to be extensional. But examples are not difficult to find in which this assumption gives rise to strange results. Let a_1 signify that John kills his wife's murderer, a_2 that he kills only other persons than his wife's murderer, and b that he does not kill anybody at all. Then $\neg a_1$ is logically equivalent with

[8] Arguably, this interpretation deviates from the common understanding of what it means to be committed to something. In ordinary parlance, commitment seems to be subject to a "committed implies can" restriction that parallels the "ought implies can" restriction. If I am committed to believe in all true mathematical statements, then this is a commitment in an entirely different sense from that in which I am committed to keep my promises and repay my loans. In a more exact analysis, such a commitment would have to be conditional on knowledge or knowability.

$a_2 \vee b$. If O is extensional, then it follows from this that $O\neg a_1$ and $O(a_2 \vee b)$ are equivalent, and in particular that

$$O\neg a_1 \rightarrow O(a_2 \vee b)$$

In words: If John ought not to kill his wife's murderer, then he ought to kill either only other persons than his wife's murderer, or no one at all. This is the revenger's paradox [31]. It can be avoided by giving up intersubstitutivity. However, that would be a far-reaching weakening of deontic logic. Therefore, just as in epistemic logic, it is customary in deontic logic to retain intersubstitutivity in spite of the problems that it can give rise to. In the case of deontic logic, one way to justify this is that the sentences that cause trouble tend to be expressed in misleading ways so that we easily overlook that they are synonymous. For instance, although the two sentences "John is obliged not to kill his wife's murderer" and "John is obliged to either kill only other persons than his wife's murderer, or no one at all" mean exactly the same thing, only the second makes it explicit that no prohibition to kill other persons than his wife's murderer is pronounced. In deontic logic, just as epistemic logic, the tradition is to accept extensionality and avoid the "intensional contexts" that give rise to trouble.

1.6.2 Logical Inference as a Means to Discover New Concepts

We can use the treatment of moral dilemmas in deontic logic as an example of how logic can be used to analyze philosophical concepts in a precise way that also gives rise to new philosophical concepts. Suppose that there is some action representable by the sentence a, such that both Oa and $O\neg a$ hold, in other words both a and not-a are morally required. This means that the dictates of the O operator cannot be completely complied with. This is the most obvious case of a moral dilemma. Indeed, moral dilemmas are often defined as situations with two conflicting obligations.

But need they be two? Suppose that someone needs to be able to reach me urgently, so that I am morally required to keep my mobile phone on. At the same time I am, for quite different reasons, obliged to be in the audience when my child performs in a school play. But members of the audience are required to keep their mobile phones turned off during the performance. Letting a denote that I have my phone on and b that I attend the performance, I am then under the three obligations Oa, Ob, and $O\neg(a\&b)$. It is easy to check that each combination of two of these three obligations is fully compatible, so there is no dilemma according to the standard definition that refers to two conflicting moral requirements. Still, the situation seems dilemmatic enough. The reason for this is of course that the combined contents of all three obligations is inconsistent. This should lead us to define moral dilemmas in terms of such combined inconsistency rather than in terms of two conflicting obligations.

For another example, suppose that I am morally required both to be in Stockholm at 10.00 a.m. (c) and to be in the neighbouring town Uppsala at 10.30 a.m. the same day (d). This is by no means logically impossible; it would indeed be practically

possible if I had access to a helicopter. But I don't. The set consisting of the two sentences c and d is not logically inconsistent. Still, this appears to express a true moral dilemma. Although the set in question is not inconsistent, it is impossible to satisfy the contents of both its elements. If we want examples like this to be regarded as moral dilemmas, then we will have to revise our definition of dilemmas so that it refers to impossibility rather than inconsistency. Realizing all elements of a finite set of sentences means the same as realizing their conjunction. We can therefore express this condition with a possibility operator $\Diamond$: A set of obligations $\{Oa_1, Oa_2, \ldots, Oa_n\}$ gives rise to a moral dilemma if and only if $\neg \Diamond (a_1 \& a_2 \& \ldots \& a_n)$, where $\Diamond$ denotes possibility.

In this way, we have generalized our original notion of a moral dilemma to the more general notion of (lack of) joint possibility (compossibility) of a set of moral obligations. This opens up for further distinctions since there are different notions of possibility. We can now speak of moral dilemmas of different types, depending on how we interpret $\Diamond$. If we interpret it as logical possibility, then we are concerned with "moral dilemmas with respect to logical possibility" which are of course much fewer than the "moral dilemmas with respect to practical possibility" that we obtain with a weaker interpretation of $\Diamond$.

Once we have formulated the issue of joint possibility of a set of moral obligations, we can generalize it further, and discuss the joint possibility of sets of norms that may contain permissions. Should we treat permissions in the same way as obligations? In other words, must the contents of a set of permissions be jointly possible in order for the set of permissions to be consistent? It is easy to show that such a requirement would be unreasonable. Just consider the set $\{Pa, P\neg a\}$, where P denotes permission and a that you take part in the weekly ceremonies of a local religious establishment. Since $a\&\neg a$ is inconsistent, such a requirement would render this set of permissions inconsistent. This is unconvincing since the very idea of religious freedom is to let us choose between such, mutually incompatible alternatives. For a set of permissions to be consistent, it seems to be sufficient that each of them, taken alone, is consistent (or possible).

Next, let us consider sets of norms that contain both obligations and permissions. There is a rather obvious way to combine the above two criteria into a single criterion that covers this more general case: Each combination of the contents of all the obligations with that of any single one of the permissions should be jointly possible. This criterion was proposed by Georg Henrik von Wright (1916–2003) ([82]; cf. [26]). It seems to work fairly well when put to test in various examples. However, calling all situations in which some permission cannot be used a "dilemma" would seem to stretch the term too far. Therefore, it may be better to use a different terminology for these cases, such as the following:

> A set of norms (obligations and permissions) is *compossible* if and only if the set consisting of the contents of all its elements is jointly possible.
>
> A set of norms (obligations and permissions) is *realizable* if and only if each subset containing all the obligations and at most one of the permissions is compossible.

A set of norms (obligations and permissions) gives rise to a *moral dilemma* if and only if the subset consisting of all its obligations is not compossible.

As we have already noted, these definitions come in different variants, depending on the standard of possibility that we apply.

We have just expressed these distinctions in natural, rather than formal, language. So what was the point of formalization in this case? The point is that it is no coincidence that these distinctions were developed by deontic logicians, rather than by moral philosophers working without the aid of a formal language. In this and many other cases, the formal language directs our attention to inference-related considerations that turn out to be helpful for the development of philosophical terminology. The usefulness of formal models is confirmed, not disconfirmed, when they give rise to distinctions that can also be expressed and used in informal philosophy.

1.6.3 Reconsidering the Formalization: Splitting Concepts

As emphasized in Sect. 1.4.4, the process of formalization should include careful consideration of whether or not terms from natural language can be treated uniformly in the formal language. However, in spite of the formalizer's best efforts, it is not uncommon that once rules of logical inference have been introduced, new problems are discovered that reveal a need to modify the original formalization. On occasions, a need for additional splitting of concepts is discovered. Consider the following two sentences, said to someone who beats a cat:

(1) "You must stop beating Mei-Yin."
(2) "You are not allowed to be cruel to animals."

(1) differs from (2) in offering a norm for only one situation, namely the present one. In contrast, (2) exemplifies the most common type of norms referring to several situations, namely normative rules.[9] This distinction is not easily extracted from deontic discourse in natural language, since most languages use the same linguistic forms for both purposes. This applies to conditional as well as non-conditional norms. Consider the following examples:

(3) "If a president from the left is elected, then rich people will have to pay more taxes."
(4) "If you bribe the head of department, then you will be permitted to take part in the extra retake."
(5) "If you borrow money, then you must pay it back."

[9]This distinction was made in [28]. Similarly, Carlos Alchourrón [1] distinguished between "a norm for a single possible circumstance (which may be the actual circumstance)" and a norm for "all possible circumstances", and David Makinson [59] distinguished between norms "in all circumstances" and norms "in present circumstances".

(6) "If you pay the exam fee at least one week in advance, then you will be permitted to take part in the extra retake."

(3) and (4) are conditional statements saying that *if* the situation satisfies (will satisfy) a certain characteristic, *then* certain actions are (will be) obligatory, respectively permitted. These statements do not report any normative rules; they only tell us what will be the case (normatively) under certain conditions. Contrastingly, (5) and (6) express rules stating that in situations satisfying the given criteria, a particular norm holds. The similarity between (4) and (6) illustrates that linguistic form does not help us to distinguish between the two types of statements. In fact, natural language provides no cue about the different types of conditionality in (4) and (6). It is our knowledge of what legal and administrative rules usually look like that makes us infer that (6) reports a permissive rule and (4) a statement about what will in fact be permitted under certain circumstances.

In order to explore the logical significance of this distinction, we can use the standard notation for conditional obligation and permission: We write $O(a \mid b)$ for "a is obligatory, given b", and similarly $P(a \mid b)$ for "a is permitted, given b".[10] Now consider the following two logical principles:

If b is true and Oa holds, then so does $O(a \mid b)$.
If b is true and Pa holds, then so does $P(a \mid b)$.

Let us first try them out on statements expressing situation-specific norms. Suppose that before the presidential election I made the statement denoted (3) above. A left-wing president is elected, and after the election it turns out that rich people are indeed required to pay more taxes. It would then be strange to claim that what I said was wrong. In particular, a rebuttal could not be based on the claim that (3) does not hold in general – the statement only referred to the specific situation. The same analysis applies, perhaps even more clearly, to statement (4). In fact, these principles apply, although perhaps less obviously, if the sentences a and b are completely unrelated. This is due to properties of "if…then …" that are unrelated to the normative component of the sentences. In a non-normative context, we would admit the following inference as valid (albeit somewhat awkward):

Xiu-xiu has a blue shirt.
Xiu-xiu knows the ancient Greek language.

If Xiu-xiu has a blue shirt, then she knows the ancient Greek language.

For the same reason we should accept the following inference:

Xiu-xiu has a blue shirt.
Xiu-xiu is permitted to read classified government documents.

If Xiu-xiu has a blue shirt, then she is permitted to read classified government documents.

[10] This notation was introduced by Bengt Hansson [27].

But this can only be true provided that we do not read the conditional statement as expressing a normative rule. From the facts that Xiu-xiu has a blue shirt and that she is permitted to read classified government documents we certainly cannot conclude that there is a rule to the effect that if she has a blue shirt, then she is permitted to read these documents. And it is easily checked that the two inference rules do not hold for rule-reporting normative statements such as (6). From the two facts that I paid the exam fee more than one week in advance and that I was permitted to take part in the exam, it does not necessarily follow that there is a rule to the effect that if one pays the fee within this time then one is allowed to take part in the exam.

Since situation-specific and rule-expressing norms are expressed in the same way in natural language, the distinction between them has often gone unnoticed. It received attention when the formal structure was put to test in logical investigations, and it turned out that they differ in what logical rules they obey. The logical differences between situation-specific and rule-expressing norms is nevertheless a good reason to make this distinction in both formal and informal philosophy, despite the fact that ordinary language does not distinguish between them.

1.6.4 Reconsidering the Formalization: Radical Reform

To illustrate how logical investigations can reveal the need for a radical reform of a formalization, we can consider the problem of so-called free-choice permissions [81, pp. 21–22]. These are permissions for someone to make a choice, for instance:

> You are allowed to marry either a man or a woman.
>
> The surgeon is permitted to take out either the patient's left or his right kidney, and transplant it to the patient's daughter.

An obvious first attempt to formalize free choice permission is to represent "or" with ordinary truth-functional disjunction ($\vee$), and this is indeed the formalization that was the starting-point of the discussion. It would then seem rather obvious that the following postulate should hold:

$$P(a \vee b) \rightarrow Pa \;\&\; Pb$$

This postulate looks innocuous when presented in connection with an example of permitted choice. However, if we also require intersubstitutivity for logically equiv-

alent sentences, then we can make derivations with highly implausible outcomes, such as the following [44, pp. 176–177]:

$P((a\&b)\vee(a\&\neg b)) \rightarrow P(a\&b)\ \&\ P(a\&\neg b)$
(a substitution instance of the postulate)

$Pa \rightarrow P(a\&b)\ \&\ P(a\&\neg b)$
(follows from the intersubstitutivity of logically equivalent sentences)

$Pa \rightarrow P(a\&b)$

The endpoint of this derivation is obviously absurd, and it gives us reason to either give up the formalization of free choice permission as $P(a \vee b)$, or else modify the framework in which the derivation took place. Following the first line, some authors have tried to solve the problem by replacing the standard permission operator P by some other operator, but such alternative operators have invariably been shown to have implausible properties [41]. The underlying problem is that all these constructions are based on the assumption that free choice permission to p or q can be represented as a property of the sentence $p \vee q$. However, if intersubstitutivity holds, then this *single sentence assumption* is not at all plausible. The reason for this is that it has the following rather obvious consequence:

> If $a \vee b$ is equivalent with $c \vee d$, then there is a free choice permission to a or b if and only if there is a free choice permission to c or d.

It is not difficult to find examples showing that this leads to absurd conclusions:

> *The vegetarian's free lunch [41]*
> In this restaurant I may have a meal with meat or a meal without meat. Therefore I may either have a meal and pay for it or have a meal and not pay for it.
>
> *Proof*
> Let m denote that you have a meal with meat, v that you have a meal without meat, and p that you pay. $((m\vee v)\&p) \vee ((m\vee v)\&\neg p)$ is equivalent with $m \vee v$. Therefore, it follows from the single sentence assumption that $((m\vee v)\&p) \vee ((m\vee v)\&\neg p)$ is (free choice) permitted if and only if $m\vee v$ is (free choice) permitted.

To sum up, in a framework with intersubstitutivity of logical equivalents, (free choice) permission to perform either a or b cannot be represented as a function of the single sentence $a\vee b$. Instead, we can treat it as a function of the two sentences a and b, i.e. as a function of two variables, not one. Similarly, (free choice) permission to perform either a, b, or c can be treated as a function of three variables, etc. Alternatively, we can treat free choice permission as a property of a set of action-describing sentences ($\{a, b\}$ respectively $\{a, b, c\}$) in these examples) [41]. In this case, logical investigations of what initially seemed to be a quite straightforward

formalization revealed the need for a rather drastic reform of that representation. In the process, we also learned something about the underlying informal notion that would otherwise not have been easy to discover.

1.7 Going Beyond Logic

The previous sections have been devoted to the use of logic in philosophy, and this for good reasons. Much philosophical subject matter is well represented by sentences, and logic provides us with powerful tools to investigate how sentences connect with each other.

But in spite of these advantages, there are no a priori grounds why logical languages should be better suited than other symbolic languages for modelling each and every subject matter studied by philosophers. In some cases, other formal approaches can capture features of the subject matter that are difficult to express in logic. It is also important to note that there is no clear demarcation between logical and "non-logical" formal methods. Arguably, much if not most of mathematics can be reconstructed in a logical framework, and conversely, logic can be seen as a branch of mathematics. But for practical purposes we can distinguish between those symbolic languages that are taught in courses and textbooks on logic and those that one has to learn elsewhere. The following subsections will briefly introduce three formal approaches of the latter category that have fairly widespread use in philosophy, namely numerical models, decision matrices, and choice functions.

1.7.1 Numbers

Numbers are ubiquitous in most of the sciences. Physicists, economists, ecologists, demographers, and scientists of almost any other discipline make frequent use of models whose variables take numerical values. Philosophy is an exception, and this for a reason that we discussed in Sect. 1.2.4: The variables that are relevant in philosophy usually cannot be correlated with empirical measurements, and therefore the most important advantage that numerical models have in other disciplines does not apply in philosophy. But nevertheless, there are cases when models involving numbers are useful in philosophy.

In value theory, it is often assumed that *value*, for instance moral value, can be expressed numerically. Moral value can then be represented by a function u, such that for each object a of evaluation, $u(a)$ is a number that represents its value. Since there is no measurement-based unit for moral value, a fictive unit is employed, often called "utile" or "util". But although the unit is elusive, the use of a numerical value function imposes a structure with considerable impact on how value is conceived. In particular, it allows us to add and multiply values. If some event has the consequences a, b, and c, we feel free to speak of their total value and

calculate it as $u(a)+u(b)+u(c)$. This, of course, is the basic structure of utilitarian moral philosophy. In contrast, deontological ethics is usually conceived in terms of the binary distinction between duties and non-duties, and therefore it has much less use for quantitative measures of value.

Standard logical models are not good at representing *time*. The reason for this is that we usually assume that for any two (non-identical) points in time, there is some third point in time that is posited between them. Such a structure is beyond elementary logic; the best way to introduce it is to employ rational (or real) numbers.

A third important area for numerical representation is *probability*. It can be introduced through a function p on event-representing propositions, such that if a is an event, then $p(a)$ is the probability of that event. Probabilities are used in epistemology, decision theory, and many other areas of philosophy. They can be given either an objective or a subjective interpretation. "Objective" probabilities represent frequencies or tendencies pertaining to events in nature. "Subjective probabilities" represent an agent's degree of belief in statements. Notably, the term "probability" should only be used about measures that have the same mathematical properties as the objective probabilities that we know from examples with coins, dice and other randomizing devices. Mathematically, this means that probabilities have to satisfy the Kolmogorov axioms.[11] It can plausibly be argued that our subjective degrees of belief should be represented by degree-of-belief functions that do not satisfy these axioms, but then they should not be called "probabilities".

In epistemology, probabilistic and logical models have complementary strengths and weaknesses. Logical models can provide us with a reasonable account of the inferential relationships among beliefs, in other words how acceptance of one belief can lead us to accept or reject some other belief. However, logical models have difficulties in representing the relations of strength among beliefs, i.e. how one belief can be stronger or weaker than another. For probabilistic models it is the other way around. They can provide good accounts of the differences in strength among beliefs, but not of the inferential connections among them [57]. Neither type of model is well suited to represent both these aspects of belief systems. That is why we need them both.

1.7.2 Decision Matrices

In a formal model of decision problems, several prominent components need to be represented. There is a set of *alternatives* that the decision-maker can choose among. In many real-life problems, the set of alternatives is open in the sense that new alternatives can be invented or discovered [32, 43]. A typical example is your decision how to spend tomorrow evening. In other decision problems, the set of alternatives is closed, so that no new alternatives can be added. Your decision how to vote in the upcoming elections will probably be an example of this. There will be

[11] See Chap. 19.

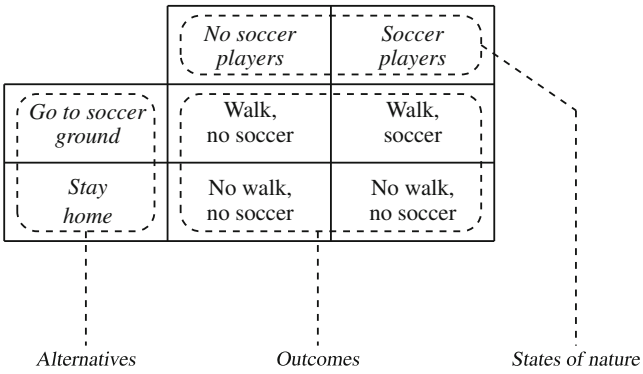

Fig. 1.7 The basic construction of a decision matrix

a limited number of alternatives (candidates or parties) that you can choose among. In decision theory, the alternative set is almost invariably assumed to be closed. The major reason for this is that formal treatment is much easier if the alternative set is closed. For the same reason, it is also commonly assumed that the alternatives are mutually exclusive, i.e., it is not possible to choose more than one of them.

The effects of a decision depend not only on the decision-maker's choice but also on various factors beyond her control. In decision theory, these extraneous factors are usually summarized into a number of cases, called *states of nature*. The states of nature include natural events but also decisions by other persons. As an example, consider a young boy, Peter, who makes up his mind whether or not to go to the local soccer ground to see if there is any soccer going on that he can join. The effect of that decision depends on whether there are any soccer players present. In decision theory, this situation can be described in terms of two states of nature, "players present" and "no players present".

The possible *outcomes* of a decision are determined by the combined effects of the chosen alternative and the state of nature that turns out to prevail. Hence, if Peter goes to the soccer ground and there are no players present, then the outcome can be summarized as "walk and no soccer". If he goes and there are players present, then the outcome is "walk and soccer". If he does not go, then the outcome is "no walk and no soccer".

The basic idea of a *decision matrix* is to tabulate alternatives against states of nature in order to show which outcome results from each combination. The decision matrix for Peter's decision is shown in Fig. 1.7. Such a matrix provides a clear presentation of the decision, but it does not contain all the information that the decision-maker needs in order to make the decision. The most important missing information concerns how the outcomes are valued and how plausible the states of affairs are.

The *values of outcomes* are usually expressed with numbers. Sometimes an empirical value measure is available, such as economic costs or gains, or the number of persons killed in an accident. But often fictitious numbers have to be used. In our

Fig. 1.8 A utility matrix

	No soccer players	*Soccer players*
Go to soccer ground	0	10
Stay home	3	3

Fig. 1.9 A probabilistic utility matrix

	.7	*.3*
Go to soccer ground	0	10
Stay home	3	3

example, we may for instance assume that Peter assigns the value 0 to walking to the soccer ground but finding no opportunity to play soccer, the value 3 to staying at home, and the value 10 to walking to the soccer ground and playing soccer there. We can then replace the basic decision matrix of Fig. 1.7 by a *utility matrix* (*payoff matrix*) in which these values take the place of the outcome descriptions, see Fig. 1.8.

Peter's decision will be influenced by how probable he believes it to be that there are any players at the soccer ground. Suppose that he takes this probability to be 0.3. Then he can replace the states of nature by probabilities, as in Fig. 1.9. This type of matrix is the starting-point in much of decision theory.

Game theory differs from decision theory in that there are two or more agents, each of whom has a set of alternatives to choose among. In game theory it is usually assumed that the outcome depends only on the decisions of the agents, so that no distinction is made between different states of nature. (This is an idealization that may of course sometimes be problematic.) In the basic game theoretical matrix for two agents, the decisions of the agents are tabulated against each other, and the outcome is determined by the combinations of their decisions. Figure 1.10 shows an example of this. Two agents, Rosa and Carmen, are going to meet for a meal. Rosa will make the food and Carmen will bring a bottle of wine. Rosa prefers red wine for meat and white wine for fish, whereas Carmen prefers white wine for all kinds of food.

Just as in decision matrices, the outcome descriptions of game matrices are often replaced by numerical values representing the values of the outcomes. In games it is important to distinguish between values for the different players. Therefore, outcome values are represented by vectors. When there are two agents, the vector $\langle x, y \rangle$ represents a situation in which the agent choosing among the rows in the

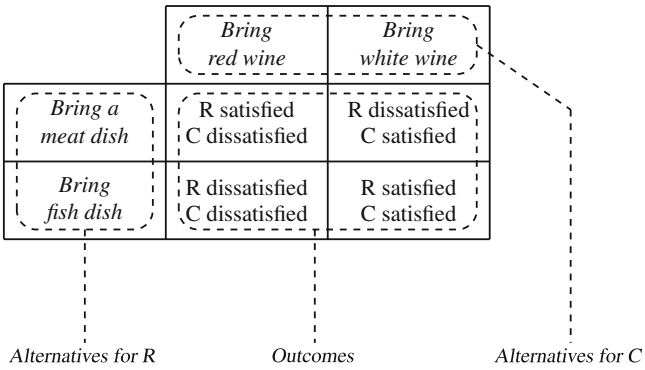

Fig. 1.10 The basic construction of a game matrix for two agents (players)

Fig. 1.11 A utility matrix for the same game as in Fig. 1.10

	Bring red wine	*Bring white wine*
Bring a meat dish	$\langle 1, 0\rangle$	$\langle 0, 1\rangle$
Bring fish dish	$\langle 0, 0\rangle$	$\langle 1, 1\rangle$

matrix assigns the value x to the outcome, whereas the other agent assigns the value y. Figure 1.11 is a utility version of the matrix in Fig. 1.10.

Both decision matrices and game matrices have turned out to be quite useful in moral and political philosophy. In particular, game matrices put focus on coordination problems that are not so easily treated in traditional logic-based models. Game matrices are also increasingly used in (social) epistemology in order to capture collective epistemic processes.

1.7.3 Choice Functions

A *choice function* is a representation of an agent's choice tendencies. If C represents your choice tendencies, then $C(\{x, y, z\}) = \{x\}$ means that if you have to choose among x, y, and z, then you will choose x. Choice functions are usually applied to sets whose elements are mutually exclusive, and they allow for ties; thus $C(\{x, y, z\}) = \{x, y\}$ means that in the choice among x, y, and z you have a tendency to choose either x or y, but you have no inclination to choose one of these rather than the other. Choice functions have an important role in decision theory,

in particular in social decision theory where a central issue is how the choices of individuals can be combined into a social choice.[12]

In addition, choice functions have turned out to be useful in several other applications, such as belief revision, non-monotonic reasoning, and the logic of conditionals. Consider an agent who initially believes in the two statements a and b, but then receives information showing that they cannot both be true. In the terminology of belief revision, the agent has to contract by the sentence $a\&b$. She must then give up either her belief in a or that in b, or both. A choice function can be used to model such choices. Virtually all belief revision models make use of choice functions, but they differ in what the choice functions are applied to: beliefs to remove, beliefs to retain, belief states that can be the outcome of the operation, etc. The application of choice functions to different types of objects in a model of a human belief system can give rise to operations of belief change with different properties [42].

1.7.4 Combinations

The use of non-logical formal tools does not mean that logic is discarded. To the contrary, the different non-logical tools are often combined with components from logic. It is for instance convenient to apply the utility function u to sentences that represent the states of affairs under evaluation, and the same applies to the probability function p. Increasingly, logicians are working with "hybrid systems" that combine logic's unsurpassed ability to represent statements and their interrelations with various non-logical formal tools that enable us to treat and rearrange these statements in ways that logic alone does not have resources for: choose among them, assign values and probabilities to them, arrange them in temporal order, etc. Such hybrid systems can sometimes make it possible to combine the advantages of two or several formal representations.

1.8 Aberrations in Formal Models

In spite of all its advantages, formalization is not always useful. In some cases it has given rise to more confusion than clarity. And even when it is useful, it is seldom if ever without problems. As pointed out in Sect. 1.2.4, a formal model is always the outcome of a trade-off between simplicity and faithfulness to the object of study. If the subject-matter is complex, then a reasonably simple model will usually have to leave out some of its philosophically relevant features.

[12] See Chap. 37.

Due to this trade-off, an uncriticizable formal model of philosophical subject matter is in practice unachievable. It will always be possible to develop a criticism that puts focus on the simplifications that are inherent in the model. However, even if such a counter-argument convincingly discloses an imperfection in the model, it does not necessarily follow that the model is unfit for use. If a problem in the model cannot be solved without substantial losses of simplicity, then it may be appropriate to continue using the model, bearing in mind its weaknesses (and perhaps supplementing it with other models that have other strengths and weaknesses). The same applies, of course, to inaccuracies in informal models and approaches in philosophy. The "adversary method" [63] in philosophy which takes any flaw in a philosophical theory as proof that the theory should be rejected in toto, is equally misguided in formal as in informal philosophy.

This is the reason why this section is called "Aberrations in formal models", rather than for instance "Faults in formal models". Depending on what we use the model for, some aberrations may be acceptable whereas others are not.

We can divide aberrations in formal models into two major types: those concerning what can be expressed in the model's language and those concerning what can be inferred in the model. Each of these types can be further subdivided depending on whether the aberration concerns an unjustified addition to what can be expressed respectively inferred, or an unjustified subtraction from it.

1.8.1 Aberrations of Expression

Almost all formal models have a conspicuous deficit in what they can express. This is mainly because in order to construct a workable formal model, the number of primitive notions has to be kept to a minimum. A few examples will be sufficient to show the rather drastic limitations in the expressive power of most formal languages. In formal value theory, only a few value concepts are represented, primarily "better" and "at least as good as", and those that can be defined from these, such as "best" and perhaps "good". In contrast, ordinary language is rich in value terms, most of which are seldom if ever included in formal accounts: "acceptable", "fairly good", "worthless", "invaluable" etc. As discussed in Sect. 1.4.4, deontic logic usually has only one concept of moral requirement (O), whereas natural language has a whole collection of prescriptive terms that differ in strengths and connotations, such as "must", "should", "ought", "has to", "duty", "obligation", etc. Epistemic logic has its focus on representations of the two terms "know" and "believe", mostly leaving out other epistemic terms such as "assume", "guess", "be convinced", "doubt", "be aware that", "have a hunch that", "suspect", etc.

Some formal languages contain superfluous expressions that do not correspond to anything meaningful that can be said about their subject matter. One way in which this comes about is through the formation rules of logical languages. If the language contains a sentential operator G for "good", then the formation rules allow us, for any sentence a, to form a statement Ga meaning "a is good". Then this will also

apply to tautologies and contradictions, and we can form sentences such as $G(b \vee \neg b)$ and $G(c\&\neg c)$ that do not seem to have a meaningful interpretation. For similar reasons, the language of deontic logic contains the sentence $O(a\&\neg a)$. It seems to express some form of moral impossibility, but it does not correspond to how we normally think about moral conundrums. Although one can say "I am obliged both to be here and not to be here", this refers to two separate obligations. ("I am obliged to be here and I am also obliged not to be here", Oa & $O\neg a$.) It does not refer to a single obligation to do something impossible ($O(a\&\neg a)$).

Such superfluous expressions can, if we so wish, be excluded from the language. Technically, this requires somewhat more complicated language formation rules than the conventional ones. In our examples, we can postulate that G and O can only be affixed to sentences that are neither logically true nor logically false. However, most logicians would be reluctant to employ language formation rules that refer to what can be logically inferred. There are good reasons to construct the language prior to, and independently of, the rules of inference. Therefore, it is much more common to retain these artefacts in the logical language, and either treat them as uninterpreted anomalies or, if possible, provide them with an interpretation that corresponds to the conditions under which they can be inferred. We can for instance treat $O(p\&\neg p)$ as an indicator of the presence of a moral conflict or dilemma.

The choice between these different ways to deal with artefacts in the logical language is largely a matter of convenience, and not very important. What is important, however, is that we do not take it for granted that all expressions in a formal language are meaningful just because they are constructed from meaningful language elements.

1.8.2 Aberrations of Inference

In some cases, the formal language does not support inferences that are reasonable and can be drawn in ordinary language. One example of this is the inference from "a is permitted" to "not-a is permitted" that we can draw in ordinary language with its bilateral notion of permission, but not in deontic logic with its unilateral notion.

But the major problem with inferences is usually the opposite one: formal models tend to support excessive inferences, i.e. inferences that are allowed by the formal system but do not correspond to any properties of that which is modelled. Arguably, most of the more problematic aberrations in formal models consist in such superfluous inference patterns. In Sect. 1.6.1 we noted that the intersubstitutivity of logically equivalent sentences produces superfluous inferences, but we also noted that for many purposes this may be an aberration that is worth its price.

A somewhat related idealization is the use of logically closed sets for various purposes in formal models, perhaps most conspicuously to represent an epistemic agent's set of beliefs. A set of sentences is logically closed if and only if everything that follows logically from it is among its elements. Hence if both a and $a \rightarrow b$

are elements of a logically closed set, then so is b. In the logic of belief revision, a logically closed set (called a "belief set") is the standard representation of an agent's beliefs. Since all mathematical truths that are expressible in a language are logical truths in that language, this means that she believes in all mathematical truths that can be expressed in the language. Such logico-mathematical omniscience is of course far beyond human capabilities. The best justification for this aberration from our actual doxastic behaviour seems to be the reinterpretation of belief sets proposed by Isaac Levi: They do not represent what an agent actually believes but what she is committed to believe. (Cf. Sect. 1.6.1 where this solution was applied to the belief operator.)

Another interesting example of excessive inferences can be taken from deontic logic. Consider the following three properties of a deontic logic:

Existence of moral dilemmas:
There are action-describing sentences a and b such that Oa & Ob, although $a\&b$ is logically inconsistent.

Agglomeration:
If Oa and Ob then $O(a\&b)$.

Necessitation:
If Oa, and a logically implies b, then Ob.

Each of these principles has immediate intuitive appeal, as can easily be confirmed with examples. But in combination they lead to an absurd conclusion. According to Existence of moral dilemmas, there are sentences a and b such that Oa & Ob and $a\&b$ is logically inconsistent. According to Agglomeration, $O(a\&b)$. Since $a\&b$ is inconsistent, it holds for any sentence c that $a\&b$ implies c. Necessitation yields Oc, and we have proved the following remarkably undesirable property:

Universal obligatoriness:
Oc

Obviously, the formal inference from Oa, Ob and the inconsistency of $a\&b$ to Oc does not correspond to how we normally reason or argue about our moral obligations. From "I ought to be at home with my children this evening" and "I ought to work all night at the office", we do not conclude "I ought to spend this evening boozing in a nightclub". Therefore, the derivation of Universal obligatoriness from three seemingly quite plausible postulates is a logical artefact that has nothing to do with the subject matter of deontic logic. Universal obligatoriness is so damaging that any system implying it will have to be rejected. Consequently, a workable system of deontic logic cannot contain all three of the principles Existence of moral dilemmas, Agglomeration, and Necessitation. The most common solution is to give up Existence of moral dilemmas. But for some purposes, such as the study of moral dilemmas, one of the other two principles will have to go instead.

If we replace necessitation by the weaker assumption of intersubstitutivity,

Intersubstitutivity of logical equivalents:
If Ox and x is logically equivalent with y, then Oy.

then the derivation of Oc (universal obligatoriness) will be blocked, but we can instead derive $O(c\&\neg c)$ (obligatory inconsistency) from Oa and Ob, given that $a\&b$ is logically inconsistent. This is also an artefact of the formal model, but as argued in Sect. 1.8.1, $O(c\&\neg c)$ does not do much damage. Arguably, it can be tolerated, and treated as an innocuous artefact of the formal system. Possibly, it can even be given a meaningful interpretation, as an indicator of the presence of an inconsistency.

1.8.3 Conclusion

In philosophy, like other disciplines, formal models are useful tools that allow us to express ideas more precisely and to probe their implications. As in other disciplines, we can only use formal models efficiently if we keep track of their strengths and weaknesses. Since all formal models are idealizations, they all have imperfections, and we should never expect to find the uniquely best formal model that will tell us the whole truth and nothing but the truth about some philosophical subject matter. But there can be no doubt that formal models are indispensable tools in philosophical investigations. Today, no philosopher can afford to be ignorant of how they can contribute to new philosophical insights.

References and Proposed Readings

Asterisks (∗) indicate recommended readings.

1. Alchourrón, C. (1993). Philosophical foundations of deontic logic and the logic of defeasible conditionals. In J.-J. C. Meyer & R. J. Wieringa (Eds.), *Deontic logic in computer science* (pp. 43–84). Chichester: John Wiley & Sons.
2. Aristotle. (1928). Prior analytics (A. J. Jenkinson, Trans.). In W. D. Ross (Ed.), *The works of Aristotle* (Vol. 1). London: Oxford University Press.
3. Aristotle. (1942). *The Nicomachean ethics* (W. D. Ross, Trans.). London: Oxford University Press.
4. Berlin, I. (1969). *Four essays on liberty*. Oxford: Oxford University Press.
5. Bobzien, S. (2002). The development of modus ponens in antiquity: From Aristotle to the 2nd century AD. *Phronesis, 47*, 359–394.
6. Brandt, R. B. (1965). The concepts of obligation and duty. *Mind, 73*, 374–393.
7. Brogan, A. P. (1919). The fundamental value universal. *Journal of Philosophy, Psychology, and Scientific Methods, 16*, 96–104.
8. Campbell, R. (1974). Real predicates and 'exists'. *Mind, 83*, 95–99.
9. Carnap, R. ([1928] 1961). *Der logische Aufbau der Welt*. Berlin: Weltkreis-Verlag.

10. Castañeda, H. N. (1981). The paradoxes of deontic logic: The simplest solution to all of them in one fell swoop. In R. Hilpinen (Ed.), *New studies in deontic logic* (pp. 37–95). Dordrecht: Reidel.
11. Chisholm, R. M., & Sosa, E. (1966). On the logic of intrinsically better. *American Philosophical Quarterly, 3*, 244–249.
12. Cook, R. T. (2002). Vagueness and mathematical precision. *Mind, 111*, 225–247.
13. Dayton, E. (1981). Two approaches to deontic logic. *Journal of Value Inquiry, 15*, 137–147.
14. Debreu, G. (1986). Theoretic models: Mathematical form and economic content. *Econometrica, 54*, 1259–1270.
15. Descartes, R. ([1644] 1905). Principia Philosophiae. In C. Adam & P. Tannery (Eds.), *Oeuvres de Descartes* (Vol. VIII). Paris: Léopold Cerf.
16. Drachmann, A. G. (1958). How Archimedes expected to move the Earth. *Centaurus, 5*, 278–282.
17. Elbourne, P. (2013). *Definite descriptions*. Oxford: Oxford University Press.
18. Forrester, M. (1975). Some remarks on obligation, permission, and supererogation. *Ethics, 85*, 219–226.
19. Frege, G. (1879). *Begriffsschrift, eine der aritmetischen nachgebildete Formelsprache des reinen Denkens*. Halle.
20. Gewirth, A. (1943). Clearness and distinctness in Descartes. *Philosophy, 18*, 17–36.
21. Gould, J. (1970). *The philosophy of Chrysippus*. Leiden: E.J. Brill.
22. Gould, J. (1980). Freedom: Triadic or tripartite? *Modern Schoolman, 58*, 47–52.
23. Green-Pedersen, N. J. (1984). *The tradition of the topics in the middle ages: The commentaries on Aristotle's and Boethius' 'topics'*. München, Wien: Philosophia Verlag.
24. ∗ Haack, S. (1978). *Philosophy of logics*. Cambridge: Cambridge University Press. [An accessible introduction to the major philosophical issues in logic.].
25. Halldén, S. (1957). *On the logic of 'better'*. Lund: Library of Theoria.
26. Hansen, J. (2014). Reasoning about permission and obligation. In S. O. Hansson (ed.) *David Makinson on classical Methods for Non-Classical Problems* (pp. 287–333) Dordrecht: Springer.
27. Hansson, B. (1969). An analysis of some deontic logics. *Noûs, 3*, 373–398.
28. Hansson, S. O. (1988). Deontic logic without misleading alethic analogies. *Logique et Analyse, 31*, 337–353, 355–370.
29. Hansson, S. O. (1990). Defining 'good' and 'bad' in terms of 'better'. *Notre Dame Journal of Formal Logic, 31*, 136–149.
30. Hansson, S. O. (1990). A formal representation of declaration-related legal relations. *Law and Philosophy, 9*, 399–416.
31. Hansson, S. O. (1991). The revenger's paradox. *Philosophical Studies, 61*, 301–305.
32. Hansson, S. O. (1996). Decision-making under great uncertainty. *Philosophy of the Social Sciences, 26*, 369–386.
33. ∗ Hansson, S. O. (2000). Formalization in philosophy. *Bulletin of Symbolic Logic, 6*, 162–175. [Short introduction to formalization, with an emphasis on its advantages and disadvantages.]
34. ∗ Hansson, S. O. (2006). How to define − A tutorial. *Princípios, Revista de Filosofia, 13*(19–20), 5–30. [Advice for the informal definitional work that should precede the construction of a formal definition of a philosophical concept.]
35. Hansson, S. O. (2006). Ideal worlds − Wishful thinking in deontic logic. *Studia Logica, 82*, 329–336.
36. Hansson, S. O. (2006). Editorial: A dialogue on logic. *Theoria, 72*, 263–268.
37. Hansson, S. O. (2006). Levi's ideals. In E. J. Olsson (Ed.), *Knowledge and inquiry. Essays on the pragmatism of Isaac Levi* (pp. 241–247). Cambridge: Cambridge University Press.
38. Hansson, S. O. (2009). The demise of modern logic? In F. Stoutland (Ed.), *Philosophical probings: von Wrights later work* (pp. 169–176). Copenhagen: Automatic Press.
39. Hansson, S. O. (2010). Editorial: Methodological pluralism in philosophy. *Theoria, 76*, 189–191.

40. Hansson, S. O. (2011). Twelve theses on the use of logic in moral philosophy. In A. Gupta & J. van Benthem (Eds.), *Logic and philosophy today* (Vol. 2, pp. 65–84). London: College Publications.
41. Hansson, S. O. (2013). Varieties of permission. In D. Gabbay, J. Horty, X. Parent, R. van der Meyden, & L. van der Torre (Eds.), *Handbook of deontic logic and normative systems* (Vol. 1, pp. 195–240). London: College Publications.
42. Hansson, S. O. (2017). *Descriptor revision. Belief change through direct choice*. Cham: Springer.
43. Hansson, S. O. (2018). Risk analysis under structural uncertainty. In T. Aven & E. Zio (Eds.), *Knowledge in risk assessment and management*. Hoboken, NJ: Wiley.
44. Hilpinen, R. (1982). Disjunctive permissions and conditionals with disjunctive antecedent. *Acta Philosophica Fennica, 35*, 175–194.
45. Hodges, W. (2009). Traditional logic, modern logic and natural language. *Journal of Philosophical Logic, 38*, 589–606.
46. Kaplan, A., & Kris, E. (1948). Esthetic ambiguity. *Philosophy and Phenomenological Research, 8*, 415–435.
47. Koj, L. (1991). Exactness and philosophy. In G. Schurz & G. J. W. Dorn (Eds.), *Advances in scientific philosophy, essays in honour of Paul Weingartner on the occasion of the 60th anniversary of his birthday* (pp. 599–609). Amsterdam: Rodopi.
48. Leibniz, G. W. ([1704] 1962). *Nouveaux Essais sur L'entendement Humain*. In *Philosophische Schriften* (Vol. 6). Berlin: Akademie-Verlag.
49. Leibniz, G. W. ([1704] 1996). *New essays on human understanding* (P. Remnant & J. Bennett, Trans.). Cambridge: Cambridge University Press.
50. Levi, I. (1977). Subjunctives, dispositions and chances. *Synthese, 34*, 423–455.
51. Levi, I. (1991). *The fixation of belief and its undoing*. Cambridge, MA: Cambridge University Press.
52. Liu, F., & Yang, W. (2010). A brief history of Chinese logic. *Journal of Indian Council of Philosophical Research, 27*, 101–123.
53. Liu, F., & Zhang, J. (2010). New perspectives on moist logic. *Journal of Chinese Philosophy, 37*, 605–611.
54. Livesey, S. J. (1986). The Oxford calculatores, quantification of qualities, and Aristotles prohibition of metabasis. *Vivarium, 24*, 50–69.
55. MacCallum, G. (1967). Negative and positive freedom. *Philosophical Review, 76*, 312–334.
56. ∗ MacFarlane, J. (2015). Logical constants. *Stanford encyclopedia of philosophy*. http://plato.stanford.edu/entries/logical-constants. [Clarifying summary of the philosophical discussions on the notion of a logical constant.]
57. Makinson, D. (1965). The paradox of the preface. *Analysis, 25*, 205–207.
58. Makinson, D. (1986). On the formal representation of rights relations. *Journal of Philosophical Logic, 15*, 403–425.
59. Makinson, D. (1999). On a fundamental problem of deontic logic. In P. McNamara & H. Prakken (Eds.), *Norms and information systems. New studies on deontic logic and computer science* (pp. 29–53). Amsterdam: IOS Press.
60. ∗ Makinson, D. (2012). *Sets, logic and maths for computing*. Dordrecht: Springer. [Although written primarily for computer scientists, this is an excellent introduction for philosophers who wish to acquaint themselves with the major formal tools used in philosophy. Exercises are included.]
61. McMullin, E. (1985). Galilean idealization. *Studies in history and philosophy of science, 16*, 247–273.
62. Mish'alani, J. K. (1969). 'Duty', 'obligation' and 'ought'. *Analysis, 30*, 33–40.
63. Moulton, J. (1983). A paradigm of philosophy: The adversary method. In S. G. Harding & M. B. Hintikka (Eds.), *Discovering reality: Feminist perspectives on epistemology, metaphysics, methodology, and philosophy of science* (pp. 149–164). Dordrecht: D. Reidel.
64. Parent, W. (1983). Recent work on the concept of liberty. In K. G. Lucey & T. R. Machan (Eds.), *Recent work in philosophy* (pp. 247–275). Totowa, N.J.: Rowman and Allanheld.

65. Peirce, C. S. (1905). What pragmatism is. *Monist, 15*, 161–181. Reprinted in Peirce, C. S. (1935) Pragmatism and pragmaticism. In C. Hartshorne & P. Weiss (Eds.), *Collected papers* (Vol. 5). Cambridge, MA: Harvard University Press.
66. Popper, K. (1935). *Logik der Forschung: zur Erkenntnistheorie der modernen Naturwissenschaft*. Wien: Julius Springer. [Translated into English as *The logic of scientific discovery*.]
67. Raz, J. (1975). Permissions and supererogation. *American Philosophical Quarterly, 12*, 161–168.
68. Robinson, R. (1971). Ought and ought not. *Philosophy, 46*, 193–202.
69. Russell, B. (1905). On denoting. *Mind, 14*, 479–493.
70. Russell, B. (1957). Mr. Strawson on referring. *Mind, 66*, 385–389.
71. Russell, B. ([1914] 1969). *Our knowledge of the external world, as a field for scientific method in philosophy*. London: Allen & Unwin.
72. Sainsbury, R. M. (1996). Concepts without boundaries. In R. Keefe & P. Smith (Eds.), *Vagueness: A reader* (pp. 251–264). Cambridge, MA: MIT Press.
73. ∗ Sainsbury, R. M. (2001). *Logical forms: An introduction to philosophical logic*. Oxford: Blackwell. [An introductory textbook in logic that pays much attention to the philosophical issues. Exercises are included.]
74. Schemmel, M. (2014). Medieval representations of change and their early modern application. *Foundations of Science, 19*, 11–34.
75. Stern, R. (2004). Does ought imply can? And did Kant think it does? *Utilitas, 16*, 42–61.
76. Strawson, P. F. (1950). On referring. *Mind, 59*, 320–334.
77. Trapp, R. W. (1978). Exaktheit in der Philosophie. *Zeitschrift für allgemeine Wissenschaftstheorie, 9*, 307–336
78. Tye, M. (1994). Vagueness: Welcome to the quicksand. *Southern Journal of Philosophy, 33*(Supplement), 1–22.
79. Tyson, P. D. (1986). Do your standards make any difference? Asymmetry in preference judgments. *Perceptual and motor skills, 63*, 1059–1066.
80. von Wright, G. H. (1963). *The logic of preference*. Edinburgh: Edinburgh University Press.
81. von Wright, G. H. (1968). An essay in deontic logic and the general theory of action. *Acta Philosophica Fennica, 21*, 1–110.
82. von Wright, G. H. (1999). Ought to be − Ought to do. In G. Meggle (Ed.), *Actions, norms, values. Discussions with Georg Henrik von Wright* (pp. 3–9). Berlin: De Gruyter.
83. Wallace, W. A. (1969). The 'Calculatores' in early sixteenth-century physics. *British Journal for the History of Science, 4*, 221–232.
84. Williamson, T. (1994). *Vagueness*. London: Routledge.
85. Wittgenstein, L. (1922). *Tractatus logico-philosophicus*, with an introduction by Bertrand Russell. New York: Humanities Press.