FROM CONCEIVABILITY TO POSSIBILITY

An Essay in Modal Epistemology

Anders Berglund

UMEÅ STUDIES IN PHILOSOPHY 8

FROM CONCEIVABILITY TO POSSIBILITY

An Essay in Modal Epistemology

Anders Berglund

Umeå 2005

Series editors: Gunnar Andersson, Ingvar Johansson, and Sten Lindström
Department of Philosophy and Linguistics
Umeå University
SE-901 87 Umeå, Sweden
ISBN 91-7305-861-0
ISSN 1650-1748
Printed in Sweden by NRA Repro AB, Umeå 2005
Distributor: Department of Philosophy and Linguistics, Umeå University,
SE-901 87 Umeå, Sweden.

Persistence

ABSTRACT

This study deals with the thesis that conceivability implies possibility. Confronted with alleged counterexamples to this thesis, some philosophers have turned to what may be called "idealized" or "more demanding" notions of conceivability. I argue that in turning to such notions, they have made the thesis useless to limited beings like us for attaining modal knowledge. However, in refusing to identify conceivability with demanding or idealized notions, we cannot maintain that conceivability always implies possibility. Essentially, there are two ways to proceed: to view conceivability as a mere guide to possibility, or to argue that the conceivability thesis is a local truth, i.e., a truth with respect to a certain class of statements. I defend the latter alternative. This class of statements employs concepts with respect to which doubt concerning the conceivability thesis is to be regarded as general skepticism, not as skepticism relating to the conceivability thesis itself.

I proceed by outlining an interpretation of strict possibility—i.e., the kind of possibility that I take the conceivability thesis to be about—according to which modal truths depend essentially on conceptual relations, as opposed to obtaining purely in virtue of properties of things themselves. Given this account, on which both ideal conceivability and strict possibility have a conceptual ground, I argue that these notions are not only coextensional but relate to one and the same property of statements. I further argue that the impossible is unimaginable, but that it is conceivable in the sense that one can misdescribe the contents of imagination.

Key words: Conceivability, conceivability arguments, possibility, modal epistemology, modal metaphysics, Descartes, Arnauld, Chalmers, Kripke, Yablo.

ACKNOWLEDGEMENTS

Central parts of the present thesis are based on my talk "What does it mean that something is conceivable?" held at *Filosofidagarna* in Stockholm, June 15, 2001. I want to thank everybody who came to listen to my talk and who contributed to the ensuing discussion. I also want to thank Peter Nilsson for valuable comments on drafts of that talk.

Philosophically, three persons have influenced me more than others. First, I want to thank Daniel Svensson. Our early discussions kindled much of the passion that enabled me to see this project through. Secondly, I want to thank Professor Joseph Almog, whose lectures and seminars have been a continuing source of inspiration. Thirdly, and foremost, I want to thank my supervisor Professor Sten Lindström for the many sessions at the blackboard and for his unswerving patience and support.

I also want to thank everybody at the Department of Philosophy and Linguistics, Umeå University. In particular, Peter Melander, Jonas Nilsson, Pär Sundström, and Peter Nilsson have provided fruitful criticism, which in many cases has led to substantial revisions. I also want to thank Per Nilsson and Anders Odenstedt for continuous support. Anders deserves special thanks for stylistic revisions of the text.

The main financial support for this work has been provided within the framework of the research project "Medvetande, Materialism och Möjlighet" ("Mind, Materialism, and Modality") financed by The Bank of Sweden Tercentenary Foundation.

On a personal note, I want to thank my son David, my parents Anita and Kennet, my sister Frida, and Monica for their support. I also want to thank the members of Umeå Kendo Club, especially Stefan Sandström.

This thesis is dedicated to David.

TABLE OF CONTENTS

CHAPTER 1

MODAL EPISTEMOLOGY

1.1 Two introductory questions

This is a thesis in modal epistemology. The qualifying term 'modal' is meant to indicate that modal epistemology is a special branch of epistemology. More familiar branches of epistemology are, for example, the epistemology of mathematics and the epistemology of empirical knowledge. Whereas the epistemology of mathematics and the epistemology of empirical knowledge are concerned with the questions if, and how, mathematical and empirical knowledge can be obtained, modal epistemology is concerned with the question if, and how, knowledge about *what is possible* and *what is necessary* can be obtained.[1]

The following questions, which I have adopted from van Inwagen (1998: 74), with some terminological changes, are central in modal epistemology:

(1) How can we know that a statement S (say, "Water $\neq H_2O$") is *possible*, or possibly *true*, when we either know that S is *false* (which is the case for "Water $\neq H_2O$") or when we do not know whether S is true *or* false?[2]

(2) How can we know that a statement S, which we know to be *true*, is also *necessarily true*?

All the problems that I intend to raise and discuss in this thesis are in one way or another derived from these questions.

As they stand, the questions (1) and (2) involve a number of presuppositions. One implicit assumption is that modal statements are either true or false, and accordingly, that there is something that makes them true or false. Another assumption is that the notion of a *statement* is intelligible. In this introductory chapter, I shall spell out the presuppositions involved in the questions above,

---

[1] The term 'modal epistemology' is used by several authors as a name of this topic. See for example Yablo 1993, van Inwagen 1998, and Casullo 2000.

[2] In what follows, I shall take "… is possible" and "… is possibly true" to be synonyms. I shall also take "… is necessary" and "… is necessarily true" to be synonyms.

and I shall explicate the various terms employed. In section 1.2, I shall discuss the notion of a *modal fact*, and the idea that there are modal facts. In section 1.3, I provide a first characterization of *modal statements* in connection to the previous discussion concerning modal facts. The explications made in sections 1.2 and 1.3 will provide a more clear demarcation of the topic of this thesis. In section 1.4, I discuss the use of modal statements in philosophical arguments. Finally, in section 1.5, I give an outline of the thesis and mention some central works in the field.

1.2 Modal facts

In this thesis, I shall assume that there are facts that make modal statements true or false, and I shall refer to these facts as *modal facts*. One can roughly distinguish between three positions regarding the nature and existence of modal facts: *eliminativism*, *reductionism*, and *primitivism*.

1.2.1 Eliminativism

According to eliminativism, there are no modal facts whatsoever. One philosopher who is extremely critical of talk about modal facts and modal properties—no matter whether these facts are taken to be facts about logic, language, or the external world—is W. V. Quine. The conclusion of Quine's criticism of modal notions, that is, the notions of possibility and necessity, is that they are unclear (see Quine 1947: 43). Moreover, Quine has systematically dismissed all attempts to clarify these notions (see Quine 1960: 195–200; 1961b [1951]: 29–30; 1998: 396). Quine is particularly critical of the notion of modality we are about to develop. According to this notion, many modal facts are facts about *the external world*, as opposed to facts about our language and the ways in which we describe things. Quine has repeatedly argued that things or states of affairs do not have modal properties independently of how we think of them or describe them (Quine 1961a: 155–6; see also Plantinga 1974: 22–4). However, Quine would also reject the idea that there are modal facts about language (for example, that some statements are "necessary" and that others are not). For Quine, the intelligibility of such a distinction between statements depends on the intelligibility of the analytic/synthetic distinction, which he rejects (see Quine 1960: 195–200; 1961b [1951]).

1.2.2 Reductionism

According to reductionism, there are modal facts, but modal facts are (reducible to) non-modal facts of some kind. There are many possible versions of reductionism: on one version, modal facts are identical to natural facts, and on another, modal facts are identical to facts about essences. Notably, the latter version seeks to reduce modal facts to facts that are equally problematic: the idea that there are *essences* in terms of which modal facts can be explained seems at least as problematic as the notion of a modal fact itself. According to reductionism, an explanation of modal concepts in non-modal terms is possible. This is the type of explanation that Quine asks for in the following passage:

> There are logicians, myself among them, to whom the ideas of modal logic [that is, the concepts of possibility and necessity] (e.g. [C. I.] Lewis's) are not intuitively clear until explained in non-modal terms. (1947: 43)

Quine's own proposal on how to clarify the concept of necessity (and, derivatively, the concept of possibility) relies on the notion of analyticity (Quine 1947). Quine's idea is that although the concept of analyticity is unsatisfactory, it definitely appears clearer than the notions of possibility and necessity (1947: 45). Accordingly, he proposes the following interpretation of '□' (the necessity operator):

(3) The result of prefixing '□' to any statement S is true iff S is analytically true.

From (3) we can derive the following interpretation of '◇' (the possibility operator):

(4) The result of prefixing '◇' to any statement S is true iff ¬S is not analytically true.

However, (3) and (4) are only applicable in propositional modal logic, where the necessity operator is applied to statements. (3) and (4) still leaves it unclear how to interpret expressions in quantified modal logic, such as

(5) $\exists x(\mathrm{Red}(x) \wedge \Diamond \mathrm{Round}(x))$.

Eventually, Quine came to reject his own proposal due to his criticism of the analytic/synthetic distinction.

Another example of a reductive explanation might be the following.[3] Some authors advocate what is sometimes called *Aristotelian essentialism*, according to which things have essential properties independently of how we describe or

[3] Quine would not approve of this explanation, due to its metaphysical implications.

think of them (for explications of the notion of essence, see section 6.3). On a standard conception, Aristotelian essentialism argues that if a thing *a* has a property *F*, it is the essence of *a* that determines whether the fact that *a* has the property *F* is a necessary or merely contingent fact. (Hence, according to Aristotelian essentialism, there are modal facts that pertain to things in the external world.) On some (but not all) formulations of Aristotelian essentialism, the concept of *essence* is not explicated in modal terms.[4] Instead, one takes it that the essential traits of a thing *a* are determined by what *kind* of thing *a* is, and this is established, it is thought, merely by considering what properties *a* has in the actual world. Further, it is argued that what is *possible* and *necessary* for *a* is determined by the essence of *a*. For example, if it is essential to *a* that *a* is a human being, then, it is argued, *a could not have been* an inanimate object. In other words, *a* is *necessarily* a human being (or animate object). In this sense, it is thought, we can explain the modal facts about *a* just in terms of what is actually the case with *a*—in terms of the essence of *a*, and in terms of what type of thing *a* (actually) is. I shall return to this idea below.

1.2.3 Primitivism

Primitivism holds that modal facts are irreducibly modal. We shall say that a fact is irreducibly modal if (i) it pertains to what could have been the case, and (ii) it cannot be reduced to facts about what is actually the case. How does primitivism fit with reductive accounts of modal facts, such as the aforementioned version of Aristotelian essentialism? The primitivist could argue as follows. Even if we grant that the following theses are true:

(i) Facts about essences are facts about what is *actually* the case

(ii) Some modal facts are *grounded* in facts about what is actually the case (such as facts about essence), and

(iii) Modal facts can be *explained* in terms of facts about what is actually the case (such as facts about essence)

this does not entail that what is possible and necessary for *a* are (also) facts *about* actuality. The idea that some modal facts are grounded in (or are dependent on) actual facts has been promoted by many philosophers (in particular by Kripke 1980). However, none of these philosophers has argued that modal facts are *identical to*, or can be *reduced to*, facts about actuality. For example, Kripke (1980) has argued that if water is in fact H_2O, this is an

[4] This view is endorsed by Almog (1991) and Fine (1994).

essential fact about water. In other words: to be H_2O is what it *is* for something to be water. Now, if water is essentially H_2O, then water is *necessarily* H_2O, and *could not have been*, for example, XYZ.[5] The latter are modal facts about water, but they are not the same fact about water as the fact that water is H_2O. Modal facts pertain to what could (and could not) have been the case with respect to water, whereas the fact that water is H_2O is a purely empirical fact which cannot be modal.[6]

1.2.4 Our preliminary position

Except for such a radical skepticism towards modal facts as Quine's, our presupposition that there are modal facts allows for most of the positions we normally would describe as "skeptical" regarding the existence of modal facts. For example, our presupposition is not incompatible with the view that modal facts and modal properties are products of the way we describe things, or facts about our language. On this view, what appear to be modal facts about the external world, are in reality modal facts about relations between concepts and descriptions in our language. But this is not to say that there are no modal facts or properties. Whereas we wish to exclude the position that there are no modal facts or properties whatsoever, we want, at least initially, to allow for positions according to which modal facts are not facts about the things in the world themselves, but rather facts about the way we think and talk, or facts about the language we use when we attribute modal properties to external things and states of affairs.

In summary, we shall adopt the following preliminary terminology. We shall say that modal facts are grounded in other facts, and we shall refer to these other facts as the *base facts* for modal facts. Base facts can be facts about traits of the external world, such as the fact that a property *F* is *essential* to an object *a*. Base facts can also be facts about logic, mathematics, language, and concepts. Such (base) facts can include:

(6) Where S and P are individual statements, the fact that S logically implies P.

(7) That every even number greater than 2 is the sum of two prime numbers.

---

[5] In accordance with Putnam 1975, we take XYZ to be a chemical compound superficially similar to water.

[6] See section 2.2 below.

(8) That the statement "All bachelors are unmarried" is *analytically* true.

(9) That water is essentially H_2O.

Modal facts that may have their grounds in (6)–(9) are, for example:

(10) Base fact: (6). Modal fact: that it is necessary that S logically implies P.

(11) Base fact: (8). Modal fact: that it is *necessarily true* that all bachelors are unmarried.

(12) Base fact: (9). Modal fact: that water is *necessarily* H_2O.

1.3 Modal statements

We define what it means that a statement is a *modal statement* in connection with our notion of a modal fact. We shall say that a modal statement is a statement that (i) contains modal terms, such as 'possible' or 'necessary', but (ii) that is not logically equivalent (in first-order logic) to any statement not containing such terms. Statements that contain modal terms but do not satisfy condition (ii) we shall call *quasi-modal statements*. An example of a quasi-modal statement is the following (which is an instance of the schema $A \vee \neg A$ of propositional logic):

(13) It is necessary that snow is white or it is not necessary that snow is white.

Modal statements are made true or false by modal facts, whereas non-modal and quasi-modal statements are made true or false by non-modal facts. To non-modal facts we count, for example, facts about the external empirical world (the total physical universe, and the history of the physical universe as a whole), mathematics, language, concepts, and logic, such as the facts in (6)–(9).

Let us proceed by asking why the various qualifications in our introductory questions (1) and (2) are made. First, with regard to question (1), why are we primarily concerned with cases in which we do *not* know that the statement in question is true? The reason is that some modal truths are derivable from non-modal truths. The following a priori principle is widely recognized as self-evident:

(14) If S is true, then S is *possibly* true.

That a statement is possible means that it could have been true, and hence, if the statement is *actually* true, then it must also be *possibly* true. It is easy to see that

any modal knowledge that can be obtained by means of applications of (14) will be unexciting. With regard to information value, to be told that a statement is possibly true when one already knows that it is *true* is like being told that $p \vee q$ is true when one already knows that p is true. We shall therefore turn our attention towards the questions (i) whether we can know that a *false* statement is nevertheless possible, and (ii) whether we can know that a statement with regard to which we have no decisive information (that is, a statement we neither know to be false nor know to be true) is possible.

Given that "S is possible" is equivalent to "¬S is not necessarily true," one can easily establish that (14) and the following a priori principle are logically equivalent:

(15) If S is necessarily true, then S is true.

Later on, we shall give more elaborate explanations of what it means for something to be necessary. For now, let us say that a statement is necessarily true if it is true in all *possible worlds*, that is, it is true relative to all possible ways in which the world could have been.

Given (15), if we know that S is necessary, we know that S is true without knowing how the world actually *is*. Conversely, one of the main problems in modal epistemology is how to effectively argue that a *true* statement is also *necessarily* true. Many statements are true but not necessarily true. For example, the statement "There are seven books on my desk" is true, but it is not necessarily true. It is reasonable to suppose that there could have been fewer books on my desk, or more. (Of course, this claim is subject to the problem we are about to discuss, namely how we can obtain such modal knowledge.)

However, some statements seem to enjoy a special status in that it is widely agreed that they are necessarily true if true, and necessarily false if false. Among these statements are the statements of logic and mathematics, such as $p \vee \neg p$, and $2 + 2 = 5$. The former statement is true, and hence necessarily true, whereas the latter is false, and hence necessarily false. The relevant principle is:

(16) If S is a mathematical or logical truth, then S is necessarily true, and ¬S is necessarily false.

In fact, we can know for any statement S of logic or mathematics that S has its truth-value by necessity *without* knowing what the actual truth-value of S is. A mathematical statement that is often used to illustrate this claim is Goldbach's conjecture, which states that every even integer $n > 2$ is the sum of two prime numbers (for example: $12 = 5 + 7$). Mathematicians have not yet found a

counterexample to Goldbach's conjecture; nor has anyone been able to prove it. Still, it is widely agreed that Goldbach's conjecture is necessarily true if true, and necessarily false if false (see, for example, Casullo 2000).

We note that (16) is not as self-evident as (14) and (15). Whereas neither (14) nor (15) depends on any particular theories about truth, possibility or necessity, (16) depends on a standard (but not universal) conception of the nature of logic and mathematics. This conception may be called into doubt. For example, a radical anti-realist with respect to logic and mathematics could hold that logic and mathematics are products of social and intellectual practices.[7] If such a radical anti-realist view were to be correct, then, since social and intellectual practices could have been different, it seems to follow that logic and mathematics could have been different, and furthermore, that the logical and mathematical truths could have been different. It thus follows from such a view that mathematical and logical truths could have been false, and if they *could have* been false, they cannot be necessarily true. Against such a view, we could object that even if our social and intellectual practices were different in ways that would make, for example, the *sentence* '$2 + 2 = 5$' true, it would still remain false that the sum of 2 plus 2 equals 5. I shall endorse the traditional view that the truths of logic and mathematics are necessarily true.

Finally, I shall clarify the notion of a *statement* that I make use of in this thesis. I shall say that a *statement* is a *sentence type* taken together with the *content* of the sentence.[8] As Kaplan (1989: 500) notes, the content of a sentence, or *what is said*, is traditionally called the *proposition* expressed by the sentence. In order to identify the proposition expressed by (communicating) a particular sentence, we sometimes need to specify the *context* of communication. A context of communication is specified by identifying *a speaker*, *a time of communication*, *a place of communication*, and *a possible world*. The contents of some sentences are independent of context. Such sentences, as '$2 + 2 = 4$', always express the same proposition in every context

---

[7] Compare Franzén 1987: 56–7: "What is essential to [mathematical] anti-realism is the view that we cannot invoke or refer to any "mathematical facts" above and beyond the facts of human social and intellectual experience and practice." Franzén refers to Wittgenstein's writings for "an initial statement of a view embodying such a radical anti-realism" (1987: 57). Wittgenstein argued that the meaning of linguistic expressions is determined by their *use* in social practice (see Wittgenstein 1958, §143ff). Mathematics is also, on Wittgenstein's view, grounded in rules determined by social practice (cf. Wittgenstein 1964 [1956], for example part V, §§28 and 35; part 1, §§33 and 61, and part II, §26; see also Anderson 1964: 481–2).

[8] One distinguishes between sentence *tokens* and sentences *types*. For example, 'Water is wet' and 'Water is wet' are two tokens of the same sentence type.

in which they are uttered. Other sentences, especially those that involve *indexicals* such as 'I', 'now', and 'here' can express different propositions in different contexts (cf. Kaplan 1989: 491). For example, if I utter 'I am hungry now' at 12.00 and Fredric utters the same sentence at 12.30, my utterance expresses the proposition that Anders is hungry at 12.00, and Fredric's utterance of the same sentence expresses the proposition that Fredric is hungry at 12.30. One further example: if I utter 'Water is XYZ' in the actual world, I express a proposition that is false in the actual world, since 'water' denotes H_2O in the actual world. On the other hand, if an inhabitant of Twin World utters the same sentence (in Twin World), he expresses a proposition that is true in Twin World, because the term 'water' as used by the speakers in Twin World refers to the substance XYZ.[9]

In the following, I shall let 'S' range over statements, and let the context determine whether I am talking about (i) the *meaningful sentence* employed in making the statement, (ii) the *proposition* expressed by the statement, or (iii) *both* the meaningful sentence and the proposition.[10] Sometimes we want to speak exclusively about properties of the meaningful sentence employed in making a certain statement, and sometimes we want to talk exclusively about the proposition expressed by the statement. For example, when I say that the statement "If all men are mortal, and Socrates is a man, then Socrates is mortal" is true in virtue of its logical form,[11] I am obviously talking about properties of the meaningful sentence employed in making that statement. (I am talking about the structural properties and linguistic content of the sentence 'If all men are mortal, and Socrates is a man, then Socrates is mortal'.) However, if I say that the statement "Water is H_2O" is necessarily true because being H_2O is essential to water, then I am concerned with the *proposition* expressed by the statement "Water is H_2O."

Sometimes we also need to talk about *both* the meaningful sentence and the proposition in order to be able to discern the differences between statements. Consider the two statements "Anders Berglund = Anders Berglund" and "I = Anders Berglund." These statements express the same proposition, that I am identical to myself. Nevertheless, they are different statements, since the first is an a priori truth and the latter is not. Whether a statement is an a priori truth or

---

[9] For more elaborate discussions of the Twin World thought-experiment and the relation between meaningful sentences and propositions, see sections 4.4.1 and 5.3.2.

[10] A similar convention is employed by Chalmers (2002a: 143).

[11] The statement "If all men are mortal, and Socrates is a man, then Socrates is mortal" is an instance of the logically valid schema $(\forall x(Fx \rightarrow Gx) \wedge Fa) \rightarrow Ga$, and is thus logically valid.

not is a fact that pertains to the meaningful sentence employed in making the statement, not to the proposition expressed by the statement. Thus, in making these observations regarding the similarities and differences between the statements, I have referred both to the meaningful sentences that express the statements, and to the proposition expressed by the statements.

From now on, we shall set aside the trivial modal knowledge we can obtain by applications of (14)–(16). We shall, as indicated by the qualifications made in (1) and (2), only be concerned with modal statements the truth of which cannot be known by means of applying the aforementioned principles. With these preliminary clarifications regarding the questions (1) and (2), I turn to the question *why* we should be concerned with finding an answer to these questions, i.e., why modal epistemology is important to philosophy in general.

1.4 Modal statements and modal arguments

Below I have listed a number of statements that have the following properties: (i) they cannot be assessed by means of (14)–(16), and (ii) they are *modal* in the sense outlined above. By anticipation, I have aimed at including modal statements that are of philosophical interest. For example, (17) is a variant of a central lemma in Descartes' famous argument for mind-body dualism in the *Meditations on First Philosophy*.[12]

(17) It is possible for me (Anders Berglund) to exist without a body.

(18) Water is necessarily H_2O.

(19) Zombies (that is, creatures that are physically and functionally identical to human beings but lack conscious experiences) are possible.

(20) It is impossible for there to be liquid wine bottles.

(21) It is possible that the number of planets is an even number.

(22) The number of planets is necessarily odd.

(23) The table at which I sit could have been two feet to the left.

---

[12] See CSM 2, Med 6. In Van Cleve's 1983 presentation of the famous argument, (17) is formulated: "It is possible that: I exist and I am unextended." In Van Cleve's presentation of the argument, this statement is derived from (i) "It is possible that: I think and nothing is extended," and (ii) "It is necessary that: if I think, then I exist." (i), in turn, is derived from (iii) "It is conceivable for me that: I think and nothing is extended," and (iv) "Whatever is conceivable for me is possible." I shall present my own interpretation of Descartes' argument in section 4.2 below.

I shall briefly indicate the origins of (18)–(23) also. (18) is a much used example of so-called *metaphysical necessity* (see Kripke 1980); (19) is a lemma in Chalmers so-called *Zombie argument* (see below, section 4.4.2); van Inwagen (1998) argues that we have non-inferential knowledge about the truth of (20) and (23); and (21) and (22) are discussed in Plantinga 1974.

Some of the statements above, such as (21), ascribe modal properties to *what is said or stated*, i.e., to *propositions*. For example, the statement "It is possible that the number of planets is an even number" ascribes the modal property of possibility to the proposition that the number of planets is an even number. Modal statements that ascribe modal properties to propositions are called *de dicto* modal statements. Whereas *de dicto* modal statements ascribe necessary, possible, or contingent *truth* to propositions, there are also modal statements that ascribe modal properties to *objects*. These modal statements are called *de re* modal statements. A *de re* modal statement asserts that a certain object possess a certain property *necessarily* or *possibly*. For example, "Water is necessarily H_2O" asserts that the substance water has the property of being H_2O by necessity. *De re* modal statements may also pertain to the way in which two objects are related to each other. For example, "The number 9 is necessarily divisible by 3" asserts, about the numbers 9 and 3, that the first is necessarily divisible by the latter. (17) and (22) above are additional examples of *de re* modal statements.[13]

Why is it philosophically important to address the question if, and how, we can come to know the truth-values of modal statements? Modal statements often play crucial roles in philosophical arguments. It does not seem to matter what area of philosophy the arguments pertain to—modal statements appear everywhere. For example, philosophers have employed modal premises in order to establish conclusions such as the following:

---

[13] It can easily be demonstrated that if a *de dicto* modal statement is true, the corresponding *de re* modal statement need not be true, and vice versa. (The following examples are loosely borrowed from Plantinga 1974 and Loux 1998.) The *de dicto* modal statement "It is necessary that if someone is sitting, he or she is sitting" is true, whereas the *de dicto* modal statement "It is necessary that the number Stephen Hawking is thinking about is an even number" is false. The former statement is true because no matter how the world would have been, it would have been true that all sitting persons are sitting. The latter statement is false because Hawking could as well have been thinking about an odd number. Now, the corresponding *de re* modal statements have opposite truth-values. The *de re* modal statement "If someone is sitting he or she is necessarily sitting" is false: anyone who is sitting could as well have been standing. The *de re* modal statement "The number Stephen Hawking is thinking about is necessarily an even number" is true: we supposed that Stephen Hawking was in fact thinking about number 2.

(24) There is a perfect being (God).

(25) The mind is not a material thing.

(26) It is impossible for there to be a necessarily existent being that is essentially omniscient, omnipotent, and morally perfect (which contradicts (24)).[14]

The difference between (24)–(26) and (17)–(23) is that (24)–(26) are the *conclusions* of the arguments they belong to, whereas (17)–(23) are *premises* of the arguments they belong to. Another difference is that (24) and (25) are *non-modal* statements. The conclusion of an argument that involves modal premises need not itself be a modal statement (but it can be, as is evident from (26)). This is what makes modal statements so useful in philosophical argumentation: you can employ considerations of what *could have been the case* and principled considerations of what is necessarily the case in order to establish a conclusion about what is *actually* the case. Consider for example the following argument (where '$\Diamond$S', for any statement S, means "possibly, S"). We begin by the modal assumption that *a* and *b* are objects such that it is possible that *a* exists although *b* does not exist, that is:

(i) $\Diamond[\exists x(x = a) \wedge \neg\exists y(y = b)]$ (Modal premise)

We assume here that '*a*' and '*b*' are *rigid designators*. In general, we say that a linguistic expression is a *rigid designator* iff it denotes the same object in each possible world.[15] Now, it is a logical truth that *a* cannot exist without *a*, that is

---

[14] These statements are adopted from van Inwagen (1998: 67–8), where he cites them as examples of conclusions of philosophical arguments involving modal premises. The premises of (24) are (i) "It is possible for there to be a perfect being," and (ii) "Necessary existence is a perfection." The premises of (25) are essentially (17) above, and "Whatever is material is essentially (necessarily) material." Finally, the premises of (26) are (i) "It is possible for there to exist vast amounts of suffering for which there is no explanation," and (ii) "If there exist an omniscient, omnipotent, and morally perfect being, there cannot also exist vast amounts of suffering for which there is no explanation."

[15] In addition to our general interpretation, one can distinguish between a number of more precise explications of the notion of a rigid designator. First of all, one can distinguish between *obstinate* and *persistent* rigid designators. A designator is *obstinately* rigid iff it denotes the same object in each possible world *w* irrespective of whether the object exists in *w* or not. By contrast, a designator is *persistently* rigid iff it denotes the same object in each possible world *w* in which the object exists, and otherwise denotes nothing. (The distinction between obstinate and persistent rigid designators is introduced by Salmon 1982: 32–4.) One further distinguishes between *strong* rigid designators and *weak* rigid designators (see Kripke 1980: 48). If a designator is rigid and denotes an entity that exists in all possible worlds, that is, a *necessary existent*, we say that the designator is strongly rigid. (For example, since the

(ii) $\neg\Diamond[\exists x(x = a) \wedge \neg\exists y(y = a)]$

Suppose now for the sake of *reductio* that *a* and *b* are identical:

(iii) $a = b$ (Assumption for *RAA*)

Using *Leibniz' law*, or *the indiscernibility of identicals*:

$$\forall x\forall y(x = y \wedge Fx \rightarrow Fy),$$

we can from (i) and (iii) obtain the following:

(iv) $\Diamond[\exists x(x = a) \wedge \neg\exists y(y = a)]$ (*Leibniz' law*, i, iii)

(ii) and (iv) yield a contradiction:

(v) $\bot$ (ii, iv)

By *reductio*, we obtain:

(vi) $a \neq b$ (*RAA*, iii–v)

From a claim about what could have been the case (that *a* could exist without *b*), we have thus come to a conclusion about what actually is the case (that *a* and *b* are distinct). When modal statements play crucial roles as premises in a philosophical argument in this way, we shall call the argument a *modal argument*.

Suppose now that in the modal argument above, *a* = my mind, and *b* = my body. Then we derive, in the manner of Descartes, the conclusion that my mind ≠ my body. The obvious way to challenge this conclusion is to challenge the truth of the crucial modal premise (i).[16] In other words, the soundness of the argument depends on the truth of the crucial modal premise (i).

At this point in the evaluation of a modal argument, our introductory questions (1) and (2) will inevitably arise. In other words: one of the main reasons why we should be concerned with modal epistemology is because we want to be able to properly assess philosophical arguments that involve modal

---

number 9 exists in all possible worlds, '9' is strongly rigid.) By contrast, if a designator is rigid and denotes a merely contingent being (that is, a being that does not exist in all possible worlds), we say that the designator is weakly rigid. (For example, Gödel does not exist in all possible worlds. Hence, 'Gödel' is a weakly rigid designator.) We note that it is only relevant to make the distinction between obstinate and persistent rigid designation in discourse about contingent beings: a rigid designator that designates a necessary being is always obstinate. There are many additional distinctions one can make with regard to the concept of a rigid designator, but these distinctions suffice to show that our general interpretation can be made more precise in a variety of ways.

[16] Given that you accept that Leibniz' law is applicable to modal properties, which supports the step from (i) and (iii) to (iv), this is the *only* way to challenge the conclusion.

premises. If we cannot provide a general answer to how (and if) we can come to know the truth-values of modal statements, such arguments cannot be properly assessed. And, as van Inwagen (1998: 67) remarks, philosophy abounds with modal arguments.

1.4.1 Modal arguments and thought-experiments

In the next section, I shall give a brief preview of the thesis. In chapter 2, I turn to the questions *if*, and *how*, we can discover the truth-values of modal statements. Before doing this, however, I want to make a brief comment on the relation between modal arguments and *thought-experiments*. Some take modal arguments and thought experiments to be the same thing. I believe that this way of using these terms is unfortunate, since there seems to be thought experiments that do not involve any modal premises. For example, consider the following thought-experiment presented in Belshaw 2000:

> *The Icebox*. You and your friend are both forty. After being involved in a car crash, you are both examined by a physician. She discovers some curious facts. One of you was born, in the normal way, forty years ago. So one of you has existed for forty years. The other was involved in some ethically dubious experiment. Though conceived sixty years ago, the other person was frozen just before birth would have occurred, preserved in that state, for twenty years, and then thawed. This person, though born forty years ago, has existed for sixty years. The experiment was a success. Freezing makes no difference to someone's lifespan. (2000: 333)

Belshaw's thought experiment is ultimately intended to support the conclusion that you would not care which one of the two persons you turn out to be. (That is, there is no reason for wanting to be the person that has existed for forty years rather than being the person who has existed for forty years plus twenty years as deep-frozen; neither is there any reason for wanting to be the latter person instead of the former.) Thought-experiments like Belshaw's could perhaps be called *intuiters*. Intuiters often involve what appear to be, in some sense, impossible assumptions. Consider, for example, the split-brain and tele-transportation scenarios described in the literature on personal identity.[17] However, I believe that, although some of the assumptions involved in the descriptions of these scenarios are impossible, this may be irrelevant to the point of constructing such scenarios: to bring out our intuitions about personal identity.

---

[17] Roger Melin (1998) discusses such cases in his thesis. See, for example, sections 2.5, 5.4, and 6.2.

My suggestion is thus that the term 'thought-experiment' has a different range of application than the term 'modal argument'. Belshaw's thought-experiment, and the examples from the literature on personal identity, could all be described as thought-experiments. However, they cannot properly be described as modal arguments.[18] In summary, an argument can be both a thought-experiment *and* a modal argument, but it may also be just a thought-experiment (such as Belshaw's argument), or it may be just a modal argument (such as the derivation from (i) to (vi) on the previous pages).

1.5 Preview of the thesis

In chapter 2, I address the question whether we can discover the truth-values of modal statements, and I discuss alternative sources of modal knowledge. These include the ideas that we can obtain modal knowledge by means of logical and conceptual analysis, intuition, imagination, and so on. In particular, many philosophers have held that *conceivability* is sometimes, or always, a reliable source of modal knowledge, whereas other philosophers have denied this. It is on the thesis that *conceivability implies possibility*, or, the *conceivability thesis*, that I focus in the following chapters.

In chapter 3, I briefly present the views on the conceivability thesis held by a number of classical philosophers. I present preliminary considerations regarding the three central analytical questions raised by the conceivability thesis: first, what should we take it to mean that something is conceivable? Secondly, what should we take "implies" to mean? Finally, what kind of possibility shall we take the conceivability thesis to attribute to conceivable things? I close the chapter by preliminary fixing a notion of possibility to work with in the subsequent chapters, that of *strict* possibility. This notion is introduced as follows. A notion N of possibility is *strict* iff the following holds with respect to N:

(27) A statement S is possible with respect to N iff the world could have been such that S was true.

It is obvious that the notion of strict possibility is a "metaphysical" notion of possibility, in that it is defined in terms of how the world could have been. The main reasons for using the term 'strict possibility' instead of the more common Kripkean 'metaphysical possibility' are (i) to dissociate the notion from well-

[18] It has been suggested to me that the cases I have mentioned all involve the *tacit* assumption that they are possible (in a wide sense). My suggestion is that, even if they *do* involve such an assumption, the fact that they do so might be irrelevant to their point.

known Kripke-inspired views about what *makes* certain statements metaphysically possible or necessary, and (ii) because I, in chapter 6, use the term 'metaphysical possibility' in a more detailed sense. What "the world could have been such that S was true" precisely means, and what determines how the world could have been, is also discussed in chapter 6.

In chapter 4, I present Descartes mind-body argument from the *Meditations* in detail (CSM 2, Med 6); Saul Kripke's mind-body argument in *Naming and Necessity* (1980); and David Chalmers' Zombie argument (1996). The focus will be on Descartes' mind-body argument. I evaluate the arguments presented by Descartes and Arnauld in their exchange pertaining to this argument, and I consider a contemporary assessment of both Descartes' argument and his reply to Arnauld's criticism due to Yablo (1990). The purpose of chapter 4 is to show how the notion of possibility introduced in chapter 3 is employed in central arguments in the philosophy of mind.

In chapter 5, I discuss possible distinctions as regards the notion of conceivability, and consider a large number of definitions of what it means that something is conceivable. The goal is to provide a comprehensive overview of the alternative options one is faced with when it comes to the task of outlining a definition of conceivability, and to evaluate what options seem most viable. To this end, I close the chapter by presenting a number of global desiderata on definitions of conceivability. I conclude that the problem with these desiderata is the following; if we want to be certain that conceivability suffices for possibility, we must put very rigorous constraints on what it takes for someone to conceive of something. On the other hand, if we attend to the epistemological problem which then arises and loosen the constraints, we can no longer be certain that conceivability implies possibility. I further conclude that these desiderata cannot be simultaneously satisfied, and I take a stand on central issues regarding the question of which desiderata should be given precedence.

In chapter 6, I raise the question of what types of facts determine the possible ways in which the world could have been. Many philosophers agree that there are objectively true modal statements, and their notions of possibility may even be extensionally equivalent. However, they need not agree on what *kinds* of fact determine the possible ways in which the world could have been. In chapter 6, I outline an account of strict modality according to which all modal truths—even "metaphysical" modal truths—have their ultimate basis in conceptual truths. As background, I discuss the relation between broadly logical and metaphysical

modality, and present a "model" for deriving metaphysical modal truths.[19] Secondly, I argue that things are identified in the actual world by means of sortal individuation, and I suggest of an interpretation of Kripke (1980) in accordance with this understanding. I further propose that the question of what it means for an object to be the *same* object throughout the space of possible worlds is determined by means of sortal considerations. As we saw in section 1.4, a designator D is *rigid* iff D denotes the same object in all possible worlds. Now, if the notion of rigid designation presupposes the notion of object sameness, then the notion of rigid designation is also dependent on sortal individuation. It is widely recognized that the "model" for deriving metaphysical modal truths can only be applied to statements that exclusively involve rigid designators. We thus distinguish between metaphysically *necessary* statements ("Water = H_2O") and metaphysically *contingent* statements ("Franklin = The inventor of bifocals") by means of the notion of a rigid designator (and the fact that the former statement, but not the latter, exclusively contain rigid designators). But then, since the notion of rigid designation itself depends on conceptual considerations (in the form of sortal identification), so does the notion of a metaphysically necessary statement.

In chapter 7, I summarize the conclusions from the previous chapters. Given the conclusions from the previous chapters, I argue that the best strategy for dealing with the standard counterexamples to the conceivability thesis is to limit the scope of the thesis to a certain set of statements. I try to characterize this set of statements by means of principled epistemological consideration, and I argue that, with respect to statements from this particular set, conceivability even for limited conceivers implies strict possibility. In other words, with respect to limited conceivers, the conceivability thesis is a local truth, restricted to a limited set of statements. Given this background, I suggest of three further theses that, together with the above, I take to provide the outlines of a modal epistemology: (i) a statement is ideally conceivable (conceivable on ideal rational reflection) iff it is strictly possible; (ii) what we *actually* imagine can always be expanded, or filled in, so as to correspond to at least one possible world (in other words, we cannot directly imagine something manifestly impossible); (iii) a modal mistake consist in *misdescribing* the content of imagination. I conclude chapter 7 by sketching a realistic position regarding the nature of concepts. The conceptualist theory of modality that I propose in

---

[19] This "model" or "template" for deriving metaphysical modal truths is (roughly) suggested of by many authors, including Kripke (1980: 3); Baldwin (2002: 18–20); Jackson (1998: 59); van Inwagen (1998: 82), and Weigel (2000: 218).

chapter 6 is, prima facie, compatible with anti-realistic positions regarding concepts, such as psychologism and conventionalism. I argue that some concepts are concepts of our own making, but that other concepts are not. According to the position I suggest of, there are concepts that are fundamental to each possible conceptual scheme. Therefore, we concede to the conventionalist that, with respect to some concepts, modal truths have their ultimate basis in human conventions. On the other hand, modal truths that involve concepts that are not of our own making cannot be subjected to a conventionalist interpretation.

To conclude the present chapter, I shall mention some central contributions to the field of modal epistemology. Since this thesis focuses on the relation between conceivability and modality, I shall only mention contributions that I take to pertain to this relation.

1.5.1 Central works in the field

In the last few years, modal epistemology has been much discussed, and it seems a safe guess that this trend will continue for some time. The revived interest is, I believe, ultimately due to Kripke's *Naming and Necessity*, which appeared over three decades ago. Arguments similar to Kripke's arguments—which themselves are reminiscent of Descartes' claims (see McGinn 1976)—appear in works which have been published in the last few years. One example is Chalmers 1996, where the so-called *Zombie argument* is presented. The Zombie argument is an argument against a particular materialist conception of mind, and it has attracted much attention in the philosophy of mind, but also in general, due to its implicit theses regarding modal knowledge. The current debate over issues regarding conceivability and modality has resulted in a number of works. In what follows, I shall provide a brief overview of them.

One of the most important contributions to the modern debate appeared before modal epistemology became fashionable. This is Stephen Yablo's "Is conceivability a guide to possibility?" (1993). Yablo defends the thesis that conceivability is a fallible guide to possibility, and he further argues that dialectical processes may resolve situations where conceivability intuitions are unclear or conflicting. Note that the thesis that conceivability is a fallible guide to possibility is a weakening of the thesis that conceivability suffices for possibility, which Yablo holds to be implausibly strong. To my mind, Yablo's paper remains the most thorough attempt to outline a substantive notion of conceivability.

Another important contribution appeared in 1998, with Peter van Inwagen's "Modal epistemology." Much will be said later about this paper, so for now, I shall only indicate the main ideas. van Inwagen's paper could be understood as a cautious comment on the practice of employing modal premises in philosophical arguments. Criticizing this practice, van Inwagen points toward the epistemological commitment entailed by possibility claims. To assert that a statement S is possible is, according to van Inwagen, Yablo, and many others, to commit oneself to the thesis that there is a whole coherent reality—a *possible world*—in which S is true. Although the assertion that S is possible involves such a commitment, van Inwagen suspects that philosophers seldom have a detailed story about such a possible world ready in order to support their modal assertions. A philosopher may confidently say that a (naturally) purple cow is possible, van Inwagen claims, although she has not devoted any thought to the question whether the relevant pigment formula, combinable with cow-DNA, is genuinely chemically possible (1998: 78).

A third central contribution is David Chalmers' "Does conceivability entail possibility?" (2002a). One can gather from the title of Chalmers' paper that he attempts to evaluate the version of the conceivability thesis that Yablo (1993) takes to be implausibly strong. In his paper, Chalmers provides a typology of definitions of what it means for something to be conceivable; he distinguishes between *positive* and *negative* conceivability, between *prima facie* and *ideal* conceivability, and so on (for these distinctions, see section 5.3 below). In terms of these distinctions, Chalmers presents two versions of the thesis that conceivability suffices for conceivability that he takes to be true. However, the versions of the thesis that conceivability suffices for possibility that Chalmers takes to be true rely on substantial idealizations of the relevant notions of conceivability. Such idealizations give rise to epistemological problems that I discuss in chapter 5 below.

In recent years, a number of doctoral dissertations on modal epistemology have also appeared. One is Katalin Balog's "Conceivability arguments" (1998), and another is Christine Weigel's "On the relationship between conceivability and possibility" (2000). I shall not have much to say about these dissertations, but I shall briefly indicate what I take the main differences between them and my work to be. Balog 1998, as well as Balog 1999, contains an interesting attempt to refute Chalmers' Zombie argument. In her discussion concerning the Zombie argument, Balog proceeds from a particular definition of what it means for something to be conceivable that she takes to be the relevant one. Alternative interpretations of conceivability are not considered. Weigel 2000 is

a defense of a version of modal realism called *Aristotelian actualism*.[20] Neither in Weigel is there any thorough attempt to problematize the notion of conceivability; instead, the focus is on modal metaphysics. She notes that conceivability intuitions can be defective due to cognitive limitations (2000: 226), but holds that conceivability intuitions are, in the ideal case, supported by sufficient empirical information (2000: 217–18; see also the passage on Kripke and empirical experiences, above). However, Weigel seems to take it for granted that the reader knows what she means by "conceivable," and there is no attempt to provide a substantial explication of the notion. It should be added that much of Weigel's discussion concerning conceivability centers on Chalmers' distinction between *primary* and *secondary* conceivability (in Weigel's terms, "1-conceivability" and "2-conceivability"; for the distinction between primary and secondary conceivability, see below, section 4.4.2). For now, we can just note that the distinction between primary and secondary conceivability is a distinction between conceiving of a proposition *p* and conceiving of an intimately related proposition *q* (but see below, section 4.4.2). Weigel's discussion regarding this distinction can perhaps be seen as an attempt to problematize the notion of conceivability. However, Chalmers' distinction between primary and secondary conceivability pertains to *what it is that you conceive of*, and not to what it means that something is conceivable.

In contrast to Balog and Weigel, the main focus of the present thesis is the concept of conceivability. My aim is to problematize this concept, and to consider alternative interpretations of what it could mean that something is conceivable. Furthermore, my interest in the concept of conceivability, and in the thesis that conceivability suffices for possibility, does not derive from any pre-existent ambition to support or reject any particular philosophical position or argument.

---

[20] The general doctrine usually referred to as *modal actualism*—the theory that only the actual world exists, and that other possible worlds are merely abstract entities—has been ascribed to Plantinga (1974), Kripke (1980), and has been defended by Stalnaker (1979). On some versions of modal actualism, the actual world is one of many possible worlds, and these other possible worlds exist—as unactualized possibilities—independently of the actual world. The difference between Weigel's Aristotelian actualism and these other versions of modal actualism is that, on Aristotelian actualism, the nature of the actual world, and the facts that obtain in the actual world, *determine* what other possible worlds there are (compare Weigel 2000: iv, 50). I believe that Kripke (1980) could be described as an "Aristotelian" actualist. This belief seems to be corroborated by Rabinowicz (2002: 14; see in particular note 10).

In the discussion that follows—a discussion concerning the questions *if*, and *how*, we can attain modal knowledge—we shall encounter other contemporary (as well as traditional) contributions to modal epistemology.

CHAPTER 2

THE SOURCES OF MODAL KNOWLEDGE

In the previous chapter, we clarified various notions involved in the following questions:

(1) How can we know that a statement S (say, "Water $\neq H_2O$") is *possible*, or *possibly true*, when we either know that S is *false* (which is the case for "Water $\neq H_2O$") or when we do not know whether S is true *or* false?

(2) How can we know that a statement S, which we know to be *true*, is also *necessary*?

In this chapter, I shall present a number of ideas on how we can come to know the truth-values of modal statements. In section 2.3, I shall present the idea that modal statements can be justified by logical and conceptual (or semantical) analysis. Some philosophers believe that the interesting questions of modal epistemology appear only *after* we have addressed the statements the truth-values of which can be settled by means of logical and conceptual analysis. For example, van Inwagen (1998) argues that our introductory questions (1) and (2) pertain to statements "whose truth values cannot be discovered *by reflection on logic and the meaning of words*" (1998: 74, emphasis added). In section 2.4, I present the idea that modal statements can be justified by having *intuitive support*. We shall see that there are very different opinions regarding the value of such intuitive support. In section 2.5, I present the idea that modal statements can be justified with reference to, in some sense, "basic" modal knowledge. This proposal incorporates the idea that the notion of basic modal knowledge is fundamental and cannot be further analyzed. In section 2.6, I present the idea that modal statements can be justified with reference to *imaginability facts*. The final proposal I consider (section 2.7) is that modal statements can be justified with reference to facts about what we can *conceive*. On this proposal, the conceivability or non-conceivability of a certain statement informs us about the modal properties of the same statement. For example, it is sometimes claimed

that *conceivability implies possibility*: if it is conceivable that the world could have been in a certain way, then the world could have been that way.

2.1 Can we discover the truth-values of modal statements?

Let us take a first step towards addressing (1) and (2) by distinguishing between two closely related questions implicit in both of them. We have already hinted that one can distinguish between the following questions:

(3) *Can* we know the truth-values of modal statements?

(4) *How* can we know the truth-values of modal statements?

In this section, I shall focus on (3). One can distinguish different ways in which philosophers have responded to this question.[1]

Some philosophers seem to argue that we can never know the truth-value of any modal statement. For example, it has been suggested that Antoine Arnauld's (1612–1694) criticism of Descartes' argument against mind-body identity can be developed into an extreme skepticism towards all modal knowledge claims whatsoever (see Yablo 1990: 159–62 and 1993: 16). This position can be referred to as *modal skepticism*. Other philosophers have maintained that we can come to know the truth-value of a comprehensive class of modal statements. This position could be called *modal dogmatism*. Descartes is perhaps the best-known modal dogmatist. We shall repeatedly return to his theory of modal knowledge in the following chapters.

It is possible to distinguish between different versions of modal dogmatism. Descartes entertains a very strong version of this doctrine (although he explicitly argues that it is possible to make mistaken modal judgments). Perhaps also Chalmers (2002a) can be described as modal dogmatist in the Cartesian sense. Other philosophers entertain more modest versions of modal dogmatism. One example is Kripke (1980), who supports some of his arguments with reference to modal intuitions. However, Kripke also provides a detailed picture of how modal thinking can go wrong. In the paper from which our initial questions were taken, van Inwagen (1998) argues that we can come to know the truth-values of modal statements such as

(5) It is impossible for there to be liquid wine bottles.

---

[1] I remind the reader that we are now concerned with modal statements that cannot be assessed simply by applying some of the following principles: (i) if S is true, then S is *possibly* true; (ii) if S is necessarily true, then S is true, or (iii) if S is a mathematical or logical truth, then S is necessarily true, and ¬S is necessarily false.

(6) The table at which I sit could have been two feet to the left.

However, he argues, we cannot come to know the truth-values of other modal statements, such as

(7) It is possible for me (Anders Berglund) to exist without a body.

According to van Inwagen, we can come to know whether modal statements concerning familiar objects in our surroundings are true or false. However, the modal statements that are philosophically interesting, such as (7), seldom concern homely objects. In so far as philosophical arguments invoke states of affairs "remote from the concerns of ordinary life" (van Inwagen 1998: 76)—and they most often do—van Inwagen is highly skeptical about the use of modal premises in philosophical argumentation. Perhaps van Inwagen's position could be called *moderate dogmatism*. It deserves to be noted that van Inwagen refers to his own position as "modal skepticism," but he admits that the name may be ill chosen. van Inwagen's decision to call his own position "modal skepticism" is perhaps due to his focus on the modal statements he believes that we *cannot* come to know the truth-values of (as opposed to a focus on the modal propositions he nevertheless believes that we *can* come to know the truth-values of). To anticipate, the different versions of modal dogmatism—especially the *moderate* versions—that we shall consider below are compatible with, or even contain, a certain amount of general skepticism towards modal knowledge claims.

We thus find that at least some philosophers have explicitly held that we can come to know the truth-values of modal statements. My aim in the remainder of this chapter is to present different proposals on *how* we can come to know the truth-values of modal statements.

2.2 How can we discover the truth-values of modal statements?

What sources of knowledge are thought to help us determine the truth-values of modal statements? Let us first mention one source of knowledge that at first appears to be of little use. One of our main sources of knowledge is sensory perception. However, empiricist philosophers have argued that empirical experiences can only tell us whether certain statements are *true* or *false*, but that such experiences cannot tell us whether a true statement is also necessarily true, or whether a false statement is possibly true (see Loux 1998: 167–8). Hence, empirical experience does not seem to be a source of modal knowledge.

However, Kripke (1980) suggests that the relation between empirical experience and modal knowledge is a bit more complicated than I have suggested here. First, Kripke argues that the following principle is a priori true:

(8) $\forall x \forall y(x = y \rightarrow \Box(x = y))$ (Necessity of identity)

(8) follows from *Leibniz' law* together with the law $\forall x \Box(x = x)$, that everything is necessarily self-identical (see Kripke 1980: 3). Now, suppose that we know that $a = b$ (where 'a' and 'b' are *rigid designators*; see section 1.2 above). For example, let '$a = b$' stand for "Phosphorus = Hesperus." When we instantiate 'a' and 'b' for 'x' and 'y' in (8), we get: $a = b \rightarrow \Box(a = b)$. This step is legitimate since 'a' and 'b' are rigid designators. By Modus Ponens we then get: $\Box(a = b)$. Then, empirical observations would (indirectly) have lead to modal knowledge: we would then know that $\Box(a = b)$. Thus, Kripke has shown that empirically grounded modal knowledge is possible. However, such modal knowledge is only *partly* empirical: the crucial modal principle (8) is a priori.[2] The empirical component is contributed by the non-modal fact that $a = b$ ("Phosphorus = Hesperus").

In this sense, empirical experience can (in part) be a source of modal knowledge after all. Yablo remarks that, given Kripke's ideas on how empirical experiences may overthrow modal convictions, even "the most conscientious and clear-headed conceiver can be refuted in a moment by the dullest observer of the passing [empirical] scene" (1990: 177). However, the knowledge that comes *strictly* from empirical experience is non-modal, and can only provide modal knowledge in conjunction with a priori principles such as (8). In sum, we can say that sensory perception cannot, *in itself*, provide modal knowledge.

However, most philosophers acknowledge sources of knowledge besides the empirical. The ideas on how to justify modal statements and modal discourse to which I now shall turn all acknowledge alternative sources of knowledge. I shall not present these ideas independently of each other. My discussion of one idea on how modal statements can be justified will normally involve references to the alternative ideas. I will also address questions that may seem unrelated to the particular idea being discussed; however, these are questions that pertain to my general project.

---

[2] For more on the notion of a priori true modal principles, see chapter 6.

2.3 Modal knowledge based on logical and conceptual analysis

It seems as if the truth-values of many modal statements can be determined by means of logical and conceptual analysis alone. Consider the following statements:

(9) If all men are mortal, and Socrates is a man, then Socrates is mortal.

(10) No one is taller than himself.

(11) Phosphorus ≠ Hesperus.

Some statements are necessary simply in virtue of their *form*. (9), for example, is, as we have already seen, an instance of a logically valid schema of first-order logic. Even if we suppose that we did not know the meaning of the terms involved in (9), we would still know that (9) is necessarily true, in virtue of being an instance of a truth of first-order logic. Now consider (10). In the notation of first-order logic, (10) reads:

(12) $\neg\exists x\,\mathrm{Taller}(x, x)$

(12) is not a first-order theorem, and thus not true in virtue of its form. However, on the standard view, there are also sentences that are necessary in virtue of the *concepts* involved. In his *The Nature of Necessity* (1974), Plantinga distinguishes between "narrow" and "broad" necessity.[3] According to Plantinga's distinction, (9) is an example of *narrow* necessity: a sentence that is necessary in virtue of its form. (10), on the other hand, is an example of a sentence that is necessary in the *broad* sense. It is clear, by the concepts involved, that (10) is necessarily true: it is impossible for someone to be taller than himself, due to the logical properties of the relation *taller than*. As Barwise and Etchemendy (1992: 54) notes in their discussion of truth-value assignments in truth-tables, there is no good mechanical method of recognizing whether a particular row in a truth-table is what they call *spurious* or not. Suppose, for example, that we build a truth-table for the following sentence:

(13) $\mathrm{Cube}(x) \wedge \mathrm{Tetrahedron}(x)$.

This truth-table will consist of four rows, reading T,T; T,F; F,T, F,F. The first row in this truth-table, T,T, is what Barwise and Etchemendy calls a *spurious* assignment of truth-values to (13): it is clear from the concepts involved that nothing can be both a cube and a tetrahedron. Barwise and Etchemendy's notion of a spurious assignment corresponds to what we would call a broadly

---

[3] The term 'wide' is sometimes employed instead of 'broad'. See for example Mason 1988: 12, and Lewis & Langford 1959 [1932]: 160–1.

logical *impossibility*, that is, a statement that is necessarily false simply in virtue of the concepts involved. "*a* is a cube and *a* is a tetrahedron" is such a statement. Further examples are "*a* is taller than *b*, *b* is taller than *c*, and *c* is taller than *a*," and "*a* is the father of *b* and *b* is the father of *a*." It is best to attempt to circumscribe the notion of broad logical necessity (possibility, impossibility) by examples. As for spurious rows in truth-tables, there is no good mechanical method of recognizing whether a particular statement is broadly logically necessary, or impossible, since this depends on the concepts involved (or, the meaning of the terms employed).

Finally, consider (11). (11) is not broadly contradictory, and is thus broadly logically possible. Nevertheless, according to Kripke, (11) is *metaphysically* impossible, that is, impossible in an even broader sense of 'possible' than that of broad logical possibility. If we agree that (11) is impossible in this sense, then there are statements that we believe to be impossible for reasons other than the broadly logical. Thus, if we agree to the claim that (11) is impossible, we also accept that there are modal statements whose truth-values cannot be determined by means of logical and conceptual analysis alone.

Someone could object here by claiming that broad logical modality is the only type of modality there is. If broad logical modality is the only type of modality there is, logical and conceptual analysis will suffice to determine the truth-values of all modal statements. For example, a philosopher who is skeptical towards modal notions could perhaps go as far as to admit that the notion of broadly logical possibility is intelligible, but hold that other notions of possibility—such as the notions of epistemic, metaphysical, and nomological possibility—are obscure.[4] Faced with this objection, one should bear in mind that many philosophers hold the opposite view. Many philosophers have argued that logical possibility (broad or narrow) has little to do with what is genuinely possible. The virtue of the concept of logical possibility is that it is fairly easy to explain what it *means* for something to be logically possible. However, it is argued, this does not mean that the notion of logical possibility has any substantial content. Mason (1988: 23) concludes his paper "Logical possibility" by saying that "we should be extremely skeptical about logical possibility," and that "we should look with care" at the supposed relation between logical possibility and possibility in an "unqualified" sense. Here it is obvious that Mason distinguishes between logical possibility and possibility in an unqualified

[4] Quine, however, is also skeptical towards broadly logical, or "analytical," modality. Such skepticism follows from his criticism of the analytic-synthetic distinction. See chapter 1.

sense. The unqualified concept of possibility is, according to Mason, a "normal, relatively familiar concept" (1988: 23). Mason briskly ends his paper thus:

> [B]eginners to philosophy are often bewildered by their teachers' assertions that it is *logically possible* to fly unaided to the moon, or that any work of artificial intelligence is pointless because its goals are *logically impossible*. I am saying that the beginners are right to be bewildered. That kind of super-science is phoney. We had better look at it closely, then stop it. (1988: 23)

van Inwagen (1998) goes even further in his criticism of the notion of logical possibility, and argues that this notion has no content. According to van Inwagen, the belief that there is such a thing as logical possibility derives from "a faulty inference from the reality of logical *impossibility*, which is real enough" (1998: 71). On the face of it, van Inwagen's view seems to be in direct opposition with the view that logical modality is all there is.

These remarks indicate that it is difficult to provide an explanation of what it means for something to be possible that is satisfactory on all counts. The concept of logical possibility has a positive feature in that it is easy to explain what it means for something to be logically possible. Furthermore, the notion of logical possibility has very little "philosophical" content: it does not presuppose any substantial philosophical theory of modality. On the other hand, one may share Mason's and van Inwagen's doubt that logical possibility has anything to do with "unqualified" possibility.

However, there is a correlation between logical modality and other more substantial notions of modality such as Mason's "unqualified" possibility or van Inwagen's possibility "*simpliciter*" (van Inwagen 1998: 72). It may be that we do not know much about the unqualified possibility of a statement if we know that this statement is logically possible. For example, it is logically possible for someone to have three heads, or to make a 300 feet long jump. But are these things *really* possible? However, if we know that a certain statement is logically *impossible*, then we know that it is impossible in any unqualified sense too, provided that we take logical possibility to be the most general type of possibility. We can thus note that skepticism towards the notion of logical possibility does not necessarily extend to the notion of logical *im*possibility.

If we bring the different views on the notion of logical modality presented here together, they are all compatible with the claim that the truth-values of some modal statements can be determined by means of logical and conceptual analysis alone. This is because we can establish, on purely logical and conceptual grounds, whether a statement is logically *impossible*, and therefore impossible in any other sense. However, we note that although it is fairly easy to explain what it means for a statement to be impossible in this sense, it is not

always easy to determine whether a statement or description *is* logically impossible. Consider the following example:

(14) $P_1 \wedge \ldots \wedge P_5 \wedge GC$

where the P_n are the five axioms of Peano arithmetic,[5] and GC is the Goldbach conjecture (which we mentioned in section 1.3). (14) is logically impossible iff there is no model of Peano's axioms in which GC is true. However, due to the complexity of this statement, it cannot, at the present state of mathematical knowledge, be determined whether the statement is impossible or not, that is, if there is a model of Peano's axioms in which GC is true.

Faced with this difficulty, we could introduce a concept of *prima facie* logical possibility, which is always practically applicable. Let us roughly distinguish between the *surface* information and the *depth* information of a meaningful sentence. Hintikka (1973: 22) writes that

> depth information is the totality of information which we can extract from a sentence by all the means that logic puts to our disposal. Surface information, on the contrary, is only that part of the total information which the sentence gives us explicitly.

If we understand "logic" in the broad sense proposed by Plantinga above, we could say that a statement or description is *prima facie* logically possible if the surface information is consistent, and we could say that a statement or description is logically possible (period) if it is consistent (period). However, the line between what should count as "surface" information and what should count as "depth" information may prove difficult to draw.

2.4 Modal knowledge based on intuition

Intuitions are supposed to be a kind of non-mediated, non-inferential insights. Some philosophers hold that *modal* intuition can provide immediate insights into the truth or falsity of modal statements. Häggqvist (1996: 120) suggests that this view is an extension of the view some philosophers have held regarding *mathematical* intuition. Many philosophers have held intuition to be the primary source of mathematical knowledge, as well as of theoretical knowledge in general. Kant, for example, held that intuition provides immediate evidence for

---

[5] The five Peano axioms (which are actually due to Dedekind) are: (i) 0 is a natural number; (ii) if n is a natural number, then n^* is a uniquely determined natural number; (iii) for all natural numbers m and n, if $m^* = n^*$, then $m = n$; (iv) for each natural number n, $n^* \neq 0$; (v) if M is a subset of $\mathbb{N}$ such that $0 \in M$ and $m^* \in M$, whenever $m \in M$, then $M = \mathbb{N}$ (Stoll 1979: 58–9).

some propositions of pure mathematics (such as propositions of geometry), and that theoretical knowledge in general is dependent on immediate intuition (see Parsons 1996: 97, Körner 1960: 26, and Kneebone 1963: 194).

Kurt Gödel is a more recent example of a philosopher and mathematician who assigns a high value to mathematical intuition. Gödel is often described as a Platonist with respect to his views on the nature of mathematics. Plato held that true mathematical statements describe relations between mind-independent objects in an abstract objective reality (Körner 1960: 16–17). Given such a conception of the nature of mathematics, intuition could be described as a kind of "intellectual vision" which allows the mathematician to perceive parts of this abstract mathematical reality. For this view, Häggqvist (1996: 120) and Parsons (1996: 95) point to Gödel (1964: 271). Gödel writes:

> [D]espite their remoteness from sense experience, we do have something like a perception also of the objects of set theory […]. I don't see any reason why we should have less confidence in this kind of perception, i.e., in mathematical intuition, than in sense perception […].

An example of a philosopher who holds that there is such a thing as *modal* intuition is Van Cleve (1983), who argues that we can intuitively "see" that a statement is possible (for more on Van Cleve's view, see section 5.3.1). Given that this kind of (non-inferential) modal intuition exists, it may be difficult to determine whether a particular judgment is based on intuition or whether it is an inference based on conceptual analysis.[6] However, by the supposed immediate character of intuitions, a way of distinguishing between what it is to intuit something and what it is to *conceive* of something suggests itself (for the notion of conceivability, see below). Häggqvist (1996: 125) writes:

> If intuition appears to be a receptive phenomenon, something coming to you, conceiving is conspicuously active. To conceive of something is an activity, something you do.

Similarly, one can distinguish between what it is to intuit something and what it is to *imagine* something. Kind (1997: 19–29) argues that one of the three main

---

[6] For example, one might feel that the truth-values of obvious modal truths, such as "A round square cannot exist" are intuitively *given*. However, van Inwagen (1998: 71) supports the claim that there cannot be round squares thus:

> A round square is logically impossible because a square must, by definition, have vertices and corners, and a round thing must, by definition, have no corners, and a round square would therefore both have and not have corners.

Here it is obvious that the insight into the impossibility of a round square is inferential. (van Inwagen's piece of reasoning is a model example of an informal proof in logic.)

characteristics of imagination is that it is an activity: imagination is like thinking, which is something you do, and not like having a belief, which is not an activity (for more on Kind's views, see section 2.6).

Some philosophers think that intuitive evidence is important in this context. Kripke, for example, holds that the intuitive content of a philosophical notion constitutes "very heavy evidence" in favor of the intelligibility of the notion in question (1980: 42). In this respect, he seems to have much in common with Gödel and Van Cleve. The notion that Kripke has in mind when making these statements is the notion of *de re* modality (see section 1.4 above). About the evidential value of intuitions in general, Kripke says: "I really don't know, in a way, what more conclusive evidence one can have about anything, ultimately speaking" (1980: 42).

Other philosophers, such as Hintikka (1999), disagree with Kripke, Gödel, and Van Cleve concerning the evidential value of intuitions. Rather, they want to describe intuitions as (heuristic) *starting points* for philosophical investigations, and they do not think that intuitive judgments are the kind of judgments on which philosophical positions should ultimately *rely*. Hintikka's skepticism concerning the evidential value of intuitions is partly because, as he says, the concept of intuition has been "watered down" in contemporary philosophy.

> When both Aristotelian forms and innate ideas became discredited, no respectable justification for intuition as a separate source of insights remained. Accordingly, by the eighteenth century the notion of "intuitive knowledge" became watered down and ended up meaning little more than immediate knowledge. (Hintikka 1999: 131)

Hintikka implies that intuitive judgments must themselves be justified, and before such justification has been provided, intuitions cannot be used in order to support other judgments (cf. 1999: 130). Other philosophers share Hintikka's diagnosis regarding the notion of intuition in contemporary philosophy. Häggqvist writes:

> In traditional usage the term "intuition" named a certain kind of (supposedly) immediate a priori cognition, or the faculty enabling one to have such knowledge. In contemporary analytical philosophy, however, this is not what is usually meant by the word. Rather, it is used to cover any quick, confident and (apparently) non-discursive inclination to assent or dissent to some non-observational statement, or a judgment which is the result of such an inclination. (1996: 119–20)

Suppose that Hintikka and Häggqvist are justified in arguing that such a conceptual change has taken place, and that this new use of the concept

"intuition" is universal. Then, of course, intuitive judgments are *not* the kind of judgments on which philosophical positions (or individual philosophical claims) should rely. What role could intuitions then legitimately play in philosophical reasoning? In her discussion of Kuhn's *The Structure of Scientific Revolutions,* Fricker (1995: 182) describes intuition as being the faculty responsible for hypothesis formation, and she identifies individual intuitive judgments with individual hypotheses. On Fricker's view, it is clear that intuition belongs to the "context of discovery," and not to the "context of justification."

Fricker's view is related to Hintikka's main point in his criticism of Kripke. Hintikka argues that instead of grounding the metaphysical framework developed in *Naming and Necessity* (1980) on modal intuitions, Kripke should instead have examined these very intuitions. Of course, if Hintikka's, Häggqvist's and Fricker's analysis is correct, even when applied to Kripke's work, then Hintikka is correct in suggesting that this is what Kripke should have done. However, it is not at all clear that Kripke has been affected by the conceptual change that the concept of intuition has undergone. Kripke's continuous appeal to intuitions is hardly an appeal to mere "hunches." Rather, I believe that it is plausible that Kripke's appeal to intuitions relies on the type of grounds Hintikka describes as outmoded and discredited. However, this remains a speculative guess. What is clear, though, is that it has not been established that Kripke is part of the conceptual change described by Hintikka and Häggqvist. Perhaps Kripke has more in common with Gödel than with his contemporaries?

What bearing does the above remarks have on modal intuitions in particular? Can modal intuitions be cited as (ultimate) *grounds* for modal judgments, or must modal intuitions themselves be (non-intuitively) grounded? In one sense, to say that modal judgments are ultimately grounded in intuition is to give up the attempt to explain *how* modal knowledge is possible. We refer to modal intuitions, but due to the nature of intuitions, being non-discursive and in one sense non-communicable, the explanation stops here. In this respect, the attempt to support modal statements by means of intuition is different from, for example, the attempt to justify modal statements by means of logical and conceptual analysis, imaginability, or conceivability (see below). If I attempt to justify a mistaken modal judgment by reciting the piece of logical or conceptual reasoning by means of which I arrived at that judgment, someone else could point to the defective steps in my reasoning. Also when it comes to imaginability and conceivability, such a dialogue is possible. Yablo (1993) argues that cases where conceivability intuitions are in conflict can (sometimes) be resolved by means of modal dialogue, and Tye's (1995) account of

imaginability is grounded in the assumption that you can determine what the object of a particular act of imagination is by means of asking the person involved about the content of his imagination.

It may thus be more fruitful to investigate the ways of supporting modal statements that admit of intersubjective scrutiny. But before turning to such proposals, which involve the notions of imaginability and conceivability, I shall mention another possible way to ground modal statements that has a lot in common with the appeal to intuitions.

2.5 "Basic" modal knowledge

van Inwagen (1998) argues that we can have non-inferential knowledge about the truth-values of basic modal statements, and he refers to such knowledge as "basic" modal knowledge. Being "non-inferential," basic modal knowledge has something in common with intuitive knowledge. Further, van Inwagen argues that

> reason—operating on a combination of "basic" modal knowledge […] and facts about how the world is put together—can expand range of our modal knowledge considerably. And where reason cannot, strictly speaking, extend the range of our modal knowledge, it can perhaps extend the range of our reasonable belief about modal matters. (1998: 70)

In other words, van Inwagen suggests that our modal reasoning should draw from a combination of the two following sources: non-inferential, basic, modal knowledge and knowledge about how the world is. van Inwagen provides what he tentatively takes to be an example of such a reasoning process (cf. 1998: 82):

> *Basic modal knowledge*: it is not possible for water to be a different physical stuff from the physical stuff it is.
>
> *Fact about how the world is put together*: water is H_2O.
>
> *Modal inference*: it is not possible for water to be some other physical stuff than H_2O.

At this point, van Inwagen raises the question of *how* we come by our basic modal knowledge. One tempting answer, van Inwagen ventures, is that we come by the basic modal knowledge that a statement S (say, "water is XYZ") is possible by constructing and intellectually examining a scenario in which S is true. That is, we, in one sense of the term, *conceive* of S. But this does not really *explain* how we know that S is possible, van Inwagen argues: what if the scenario we come up with is itself impossible?

It seems as if these requests for prior assurances of possibility lead to some kind of circularity or infinite regress. However, van Inwagen argues that there is no need to go down that path, because we do not need an account of *how* we, ultimately, know that some modal statements are true and others false—we can know the truth-value of a modal statement without knowing *how* we know the truth-values of such statements. In fact, van Inwagen points out that he regards modal knowledge—knowledge that he nevertheless argues that we possess—as "deeply mysterious" (1998: 73–4). In this respect, van Inwagen takes a different course than the philosophers who have argued that the (intellectual) construction of a scenario in which S is true *is* the ultimate ground on which possibility judgments should be based. van Inwagen thus seems to have much in common with philosophers who appeal to intuition.

However, van Inwagen's suggestion that modal knowledge originates from *a combination* of different types of knowledge is similar to some accounts that do involve appeals to conceivability. In her thesis, Weigel (2000: 217–18) argues that modal knowledge is possible, and that it derives from the three following sources: (i) *a priori principles of possibility* (Weigel (2000: 218) mentions: "the chemical components of a chemical kind are the very constitution of the chemical kind"); (ii) *empirical facts about the actual world*, and (iii) *conceivability*. The difference between van Inwagen and Weigel is that Weigel appeals to *a priori principles of possibility* where van Inwagen appeals to *basic modal knowledge*. Now, the instance of basic modal knowledge that van Inwagen cites comes, loosely speaking, close to being an "a priori principle of possibility" in Weigel's sense: "for all *x*, it is not possible for *x* to be made of a different physical stuff than the physical stuff it is actually made of" (Weigel 2000: 218). However, van Inwagen also cites examples of what he takes to be basic modal knowledge that are obviously a posteriori: "John F. Kennedy could have died of natural causes" (1998: 70). It is easy to imagine a course of events (rather similar to the actual one) in which John F. Kennedy died of natural causes. So this seems to be a typical example of modal knowledge based on conceivability.

I should also mention that there at first seems as if van Inwagen's discussion is incoherent on one point.[7] Towards the end of his paper, van Inwagen presents Yablo's (1993) account of how modal knowledge is possible. Yablo's account involves the idea that a proposition or statement is *conceivable* for a conceiver iff the conceiver can imagine a world such that the conceiver takes the

[7] My discussion here will go beyond the notion of basic modal knowledge, and anticipate discussions to follow regarding the concepts of conceivability and imaginability.

proposition or statement to be true in that world (1993: 29). For example, "Water is XYZ" is conceivable for me if I can imagine a world such that I take "Water is XYZ" to be true in that world. Yablo's idea is that if a statement is conceivable in this sense, then this is prima facie evidence that the statement is possible. Now, van Inwagen endorses Yablo's account. But then, why is van Inwagen so negative towards the idea that we can come by the basic modal knowledge that S by constructing and intellectually examining a scenario in which S is true? At least prima facie, the latter idea and Yablo's idea seem to have a lot in common.

The apparent incoherence can be dispelled by making a distinction between imagining *scenarios* and imagining *possible worlds*. In a concrete sense, there is no such thing as imagining possible worlds directly: possible worlds are far too complex for us to be able to imagine them. As Yablo notes, our imaginings are usually not fully determinate in themselves, and they are uninformative about states of affairs outside the immediate focus of the imagination.[8]

Nevertheless, we do imagine some kind of situations, let alone farther less complete situations than full possible worlds. We shall call such an imagined situation a *scenario*. It is reasonable to believe that imagined scenarios are *never* complete enough to correspond to one single possible world. Consider the following passage by van Inwagen:

> Suppose, for example, that I attempt to imagine a world in which George Bush was re-elected in 1992. It is unlikely that my effort will in any way involve the state of higher education in Pakistan in the year 1957—although, of course, in every world in which Bush won [the] re-election, something or other was the case as regards higher education in Pakistan in 1957. (1998: 77)

Although van Inwagen speaks about his attempt to imagine a *possible world*, I believe that we can describe him as attempting to imagine a possible world *by means of imagining a scenario* (since he acknowledges the fact that his imagination is incomplete in the sense discussed).[9] As van Inwagen notes, for each possible world in which Bush won the re-election, something or other was

---

[8] Yablo 1993: 28–9. For examples illustrating the vagueness of imagination, see the next section.

[9] The same can be said concerning Yablo's account. The reason why Yablo wants to speak about imagining possible worlds rather than scenarios is because, as he says, "[o]n the usual theory, propositions have truth values not in limited situations [that is, in scenarios], but in the complete situations I have identified with possible worlds" (1993: 29). He goes on to note, as does van Inwagen, that "in principle, it should be possible to imagine a *p*-verifying world [where *p* is a proposition] while leaving matters visibly irrelevant to *p*'s truth value unspecified" (p. 29).

the case as regarding the state of higher education in Pakistan 1957. In one possible world where Bush won the re-election, there was no higher education in Pakistan 1957, and in another there was, and so on. Therefore, an imagined scenario (say, the scenario where Bush won the re election) will correspond to a *number* of possible worlds (the world where there was no higher education in Pakistan 1957; the world where there was; and so forth).

In summary, even if we cannot imagine possible worlds directly, we could still claim that we can imagine possible worlds, but that we imagine them *via* scenarios. In other words, we can legitimately describe our imaginings as being of possible worlds in a transferred sense.

2.6 Imagination and modal knowledge

Imagination is often involved in modal knowledge claims. Above (section 2.5), we saw Yablo argue that the *imaginability* of a possible world that one takes to be a world in which S is true constitutes prima facie evidence for the possibility of S. We have also made extensive use of the concept of imaginability in the previous section. The foremost problem with appeals to imagination is that the term 'imagination' is used to cover a variety of activities. Here, we shall not, as is sometimes done, use 'imagination' as a term covering intuition, conceptual analysis, and conceivability. Instead, we shall ask whether there are any characteristic features of imagination that would help us distinguish imagination from intuition, conceptual analysis, and conceivability.

We have already argued that there is a difference between imagination and intuition in that imagination seems to be an *activity* (at least according to Kind 1997), whereas intuition is receptive. But it may be argued that this is not the most important difference between them—or between imagination and conceptual analysis or conceivability for that matter. What seems to be particular to imagination is that at least one form of imagination has a *phenomenological* mode of presentation. Many philosophers have argued that this mode of presentation is caused by the presence of *mental imagery*: when we imagine things, it is held, we do this by means of *mental images*. There are many problems with this view. Some are ontological: in what sense of "exist" do mental images exist? Surely, there are no physical images present inside our heads when we imagine things. Another related problem is that the notion of mental images may appear unscientific. In what way could the idea that there are mental images fit within an otherwise scientific conception of the world?

The latter of these problems is probably what led many behaviorists to reject the idea of mental images. A chief exponent of the view that there are no mental

images is Ryle. See for example Ryle 1949: 245, where he argues that the answer to the question "Where do the things and happenings exist which people imagine existing?" is that they do not exist anywhere. Nowadays, however, philosophers are less hostile towards the notion of mental images. As Tye remarks, it has become respectable again "to talk of people having mental images and inquire into their role in cognition" (1991: xi).

Some philosophers even argue that mental images play an *essential* role in imagination. Kind (1997) attempts to show that there are in fact mental images present when we imagine. She further argues that if it can be shown beyond doubt that imagination has special traits that only can be accounted for in terms of imagery and mental images, then such a theory of imagination should be endorsed. According to Kind (1997: 19–24), any acceptable theory of imagination must give an account of the following three characteristics of it:

> *Imagination is directed*. To imagine something is to be directed towards that object or situation.
>
> *Imagination is active*. When I think, reason, or make a decision, this is something that I *do*. In this respect imagination is like thinking, which is something we do—and not like having a belief, which is not an activity.
>
> *Imagination has a phenomenology*. "Generally," says Kind, "when we attribute a phenomenological character to a mental attitude, the explanation for this stems from the fact that the occurrence of the mental attitude involves a state with phenomenological character, that is, a sensation" (1997: 23–4). There is *something it is like* to imagine something.

Kind ultimately concludes that the fact that imagination has a phenomenology cannot be explained unless one allows for an explanation in terms of mental imagery and mental images. Kind admits that you could explain how imagination is directed and account for the fact that it is active *without* reference to mental imagery. However, she argues, other theories will not be able to explain the fact that imagination has a phenomenology (1997: 31).

It may seem as if Kind's argument begs the question. (There is, though, more to this argument than I have mentioned here.) However, I shall nevertheless accept the thesis that imagination involves, and proceeds by way of, mental imagery. If this thesis is rejected, it will be difficult to explain what imagination consists in, and it will also be difficult to explain in what way it is different from conception, conceptual analysis, and other cognitive activities. In other words:

if we believe that imagining is something essentially different from these other activities, what this difference consists in is best explained on the assumption that imagination essentially involves mental imagery, and that the other activities do not.

The contemporary debate regarding mental images is not concerned with the question whether there *are* mental images. Rather, many participants in this debate have accepted that there are mental images. Instead, the debate centers on the question of what the nature of a mental image is. Both Tye (1991) and Block (1981) imply that this discussion can, in the main, be described as a dispute between two positions regarding the nature of mental images. According to *pictorialists*, mental images are like real images, with the exception that they do not "exist" in the literal sense of the word. According to this position, mental images represent things in the same way as real images do. In this sense, mental images are not unlike photographs. By contrast, *descriptionalists* hold that mental images have more in common with descriptions. The issue, says Dennett (1981b: 88), is whether mental imagery has more properties in common with pictures than with sentences, or if it is the other way around.

There are plausible arguments for the descriptionalist view. Dennett (1981a) observes that imagination can be vague in a variety of ways. For example, if I imagine a man with a wooden leg, some details may be left out. The color of the man's hair may be vague, but this does not entail that it is *colored* vague, says Dennett—the color of the hair is simply not "mentioned" in my imagination.[10] In this respect, imagining is more similar to *describing* than to *depicting*, Dennett argues:

> If I write down a description of a person, it would be absurd for anyone to say that my description cannot fail to mention whether or not the man is wearing a hat. My description can be as brief and undetailed as I like. (1981a: 54)

A picture of a man, on the other hand, given that it is of the *whole* man, must indicate whether he has a hat or not. An argument for the claim that mental images have more in common with sentences and descriptions than with pictures is what we may call the *striped tiger argument*.[11] Since we cannot "read off" the

---

[10] See Dennett 1981a: 54. Other authors make the same observation. See Yablo 1993: 27–8, and Tidman 1994: 300.

[11] Dennett 1981a: 55:

> If seeing or imagining is having a mental image, then the image of that tiger—obeying the rules of images in general—must reveal a definite number of stripes showing, and one should be able to pin this down with such questions as 'more

number of stripes on an imagined tiger, the imagination of the tiger is not a picture of a tiger; one *would* be able to "read off" the number of stripes from, for example, a painting or photograph of a tiger.

Similarly, the pictorialist attempts to demonstrate that mental images have more in common with pictures than with descriptions. Fodor (1981) aims to show that mental images can have none of the properties we commonly associate with descriptions, such as being true or false, carry a distinguishable and specific descriptive content, or being capable of referring to whatever they resemble. The idea that thoughts are images may be understood in two ways, says Fodor:

> On the one hand, the proposal might be that we should identify having an image with thinking *of* something, and, on the other, it might be that we should identify having an image with thinking *that* something. (1981: 67)

This distinction is also discussed by Yablo (1993: 27). Yablo distinguishes between what he calls *objectual* and *propositional imagination*. Objectual imagination is directed towards objects, in a wide sense of "objects" (the object does not have to be real or even fully determinate—see the "striped tiger argument" above). For example, when I imagine a tiger, my imagination is objectual. Propositional imagination, on the other hand, is not exclusively directed towards objects. When I imagine that there is a tiger in the hallway, says Yablo (1993: 27), I imagine a tiger (objectual component), and I imagine *that* it is behind the curtain (propositional component). Fodor and Yablo both observe that only propositional imaginations are capable of being true or false. Hence, at least objectual imaginations are not like descriptions of objects or states of affairs, since a description of an object or a state of affairs can be true or false (about the object or state of affairs), whereas objectual imaginations

---

than ten?,' 'less than twenty?' If, however, seeing or imagining has a descriptional character, the questions need have no definite answer. Unlike a snapshot of a tiger, a description of a tiger need not go into the number of stripes at all; 'numerous stripes' may be all the description says.

Yablo 1993: 27:

> When I imagine a tiger I imagine it as possessed by some determinate striping—what else?—but there need be no determinate striping such that I imagine my tiger as striped like *that*; the content of my imagination is satisfiable by *variously* striped tigers, but not by tigers of *no* determinate striping. Likewise for situations: even if there is much about my tiger-situation that I leave unspecified as irrelevant […] still I think of these things as fully definite in the situation itself.

cannot. There is no sense in saying that my (objectual) imagination of a tiger is "true" or "false."

Yablo's account of imagination differs from the pictorialist account, and perhaps also from the descriptionalist account, with respect to one important point. As Yablo points out:

> Some philosophers use "imagine" so that imagining a thing is *imaging* it, that is, conjuring up an appropriate sensory presentation. I do *not* require a sensory-like image for imagining, and certainly not a distinct such image for distinct imaginings. (1993: 27)

Yablo's comment, taken together with our previous discussion, seems to imply that there are *two kinds* of imagination: one that involves imagery, and another that does not. However, I believe that notions of imagination according to which imagination does not involve any sensory presentations—or *representations*—tend to blur the distinction between imagination on the one hand and intuitions, conceptions, and so on, on the other. According to the picture developed here, "imagination without sensory (re)presentation" is a misnomer, and the corresponding mental acts—whatever they are—belong under a different heading.

We shall not take a stand on the debate that has come to be known as *the imagery debate*, that pertains to the question whether mental images are more like pictures than descriptions, or vice versa. (However, we have objected to notions of imagination according to which imagination does not, or *need* not, involve sensory-like presentations.) Even if we refrain from taking a stand in the general question in the imagery debate, some of the descriptionalist ideas seem to fit together with our earlier ideas on what it means to imagine a scenario (see section 2.5). We have pointed out that in a concrete sense, complete individual possible worlds cannot be imagined—possible worlds are by far too complex for this to occur. However, in the same way as my imagination of a tiger may be uninformative with regard to the exact number of stripes on the tiger's back, my imagination of a possible world in which S can be equally uninformative on matters that are irrelevant to the truth of S in that world (cf. Yablo 1993: 28–29, and van Inwagen 1998: 77). This is what we have called *imagining a scenario*, and we say that a possible world in which S is true is imagined *via* the scenario. Thus, when I imagine a tiger in the hallway, I am imagining a possible world—a type of world—in which there is a tiger in the hallway. However, in virtue of the fact that my imagined tiger does not have a determinate number of stripes on its back, I am imagining a *scenario* that is satisfiable by a number of different possible worlds.

In summary, imagination is sometimes a factor in modal knowledge claims, since many authors take *the imaginability of an* S-*world* to provide evidence for the possibility of S (see Chalmers 2002a; Yablo 1993; van Inwagen 1998). This seems to imply that considerations of imaginability always come in on the possible-worlds level. (In our terms, since possible worlds are imagined *via* scenarios, considerations of imaginability come in on the *scenario*-level.) Although this is a correct description of the role imagination usually has in contemporary discussions of modal epistemology, some authors discuss, for example, the question whether the imaginability of *states of affairs* should be taken to provide modal knowledge. For example, we shall see below how Tidman (1994) and Kind (1997) reject imaginability as a source of modal knowledge, since we seem to be able to imagine impossible states of affairs (see section 5.5).

We have pointed out that there are differences between, for example, intuition on the one hand and imagination and conceivability on the other (intuition is receptive whereas imagination and conceivability are active; furthermore, imagination essentially involves representations, whereas, at least in one sense of 'intuition', intuition does not). Given the nature of logical and conceptual analysis, it is fairly certain that there are differences between that activity on the one hand, and intuition and imagination on the other. As we finally turn to the claim that there is a relation between *conceivability* and modal knowledge, we may ask what the difference between imagination and conceivability is. However, when it comes to the relation between imagination and conceivability, things are not so clear. First of all, many authors use 'conceivable' and 'imaginable' in a way that seems to suggest that these terms are interchangeable. I believe that in some senses of the two terms, they *are* interchangeable. (Here I am particularly thinking about certain wide senses of 'imaginable'.) Consider, for example, Yablo's notion of "propositional imagination." When propositional imagination is accompanied by objectual imagination, one (roughly) first imagines some object (this is the objectual component) and then one imagines that something is true about that object, or that the object has some certain property (this is the propositional component). Now, let the object of the objectual imagination be a certain *scenario* (say, a tiger in the hallway with some of its surroundings) and let the propositional part of the imagination be that some proposition or statement is true in that scenario (e.g., "There is a tiger in the hallway"). At least on a rough approximation of Yablo's own notion of conceivability, the latter instance of *propositional imagining* can as well be described as an instance of *conceiving* (cf. Yablo 1993: 27–29). In other words,

it remains an open question whether there are any substantial differences between propositional imagination and conceivability.

However, it is easy to find senses of 'conceivable' and 'imaginable' according to which there is a difference between imaginability and conceivability. For example, we argued that imagination involves sensory representations. There are definitions of what it means that something is conceivable which make no reference to representations whatsoever. Consider the following definitions (Balog 1999 and Reid 1994 [1863]):

(15) A statement S is conceivable if it is consistent with the totality of conceptual truths, that is, if ¬S is not a conceptual truth. (Balog)

(16) To conceive of a proposition is to understand distinctly its meaning. (Reid)

(17) To conceive of a proposition is to give some degree of assent to it, however small. (Reid)

It is obvious that according to (15), conceivability has nothing to do with the mental or cognitive faculties of any conceiver, let alone with imagination. According to (16) and (17), it is obvious that conceivability is a matter of cognition and reasoning rather than representation.[12]

2.7 Conceivability and modal knowledge

I shall begin this section with a brief description of a contemporary argument in the philosophy of mind. The *Zombie argument* (Chalmers 1996) is an attempt to refute the materialist claim that no possible world *w* can be identical to our world with respect to physical facts without being such that all the (positive) mental facts about our world also hold in *w* (cf. Chalmers 1996: 123; Hill and McLaughlin 1999: 445). First, we define a *zombie world* as a world that is an exact physical duplicate of our world, but where there are no conscious experiences. The materialist claim implies that there are no zombie worlds. Now, assume, on the contrary, that there is a zombie world *w*. Then each one of us would have a physical duplicate in *w*. According to the materialist thesis, our duplicates would have a capacity for experiencing pain just as we have. However, this contradicts the assumption that our duplicates are zombies (that is, that they lack conscious experiences). Hence, according to materialism, no possible world is a zombie world.

---

[12] Balog's and Reid's definitions are three examples of the many definitions of what it means for something to be conceivable that we shall be concerned with in chapter 5.

In rough outline, Chalmers' Zombie argument involves the following claims. First, Chalmers argues that a zombie world is conceivable by means of "ideal rational reflection" on the concepts employed in the description of a zombie world.[13] That is, a description of a zombie world does not involve any logical or conceptual contradiction. From this, Chalmers eventually concludes, via an elaborate argument we shall consider in section 4.4.2 below, that a zombie world is possible. Hence, the materialist thesis must be false: there is a possible world that is exactly like our world with respect to physical facts without being such that all positive mental facts about our world also hold.

The zombie argument obviously turns on the crucial inference from the *conceivability* of a zombie world to the *possibility* of such a world. In general, the possibility of a zombie world is established by an application of *the conceivability thesis*:

> *The conceivability thesis*: conceivability implies possibility.

Chalmers' argument does not make *direct* use of the conceivability thesis. The first part of his argument merely establishes that (the description of) a zombie world is conceptually consistent, or, broadly logically possible. However, in the latter part of the argument, Chalmers argues that there is no distinction between logical and metaphysical possibility on the level of worlds. (For details, see section 4.4.2) In summary, we can say that the zombie argument starts from a claim about *conceivability*, proceeds from there to a claim about *possibility*, and from there, to a substantial conclusion about the nature of consciousness.

Not surprisingly, Chalmers' Zombie argument has been subjected to extensive criticism. Most of the criticism has focused on Chalmers' claims about conceivability and possibility. Some, like Nagel (1998), seem to argue that a zombie world is not *genuinely conceivable*:

> [T]he powerful intuition that it is conceivable that an intact and normally functioning physical human organism could be a completely unconscious zombie is an illusion—due to limitations of our own understanding. Nevertheless those limitations are real. (1998: 342)

Others, like Loar (1999: 466–7) have argued that the zombie argument only establishes that a zombie world is possible in a weak sense of the word "possible."

---

[13] See Chalmers 1996: 68, 98, 100, and 131. See further Chalmers 1999b: 436. For the notion of conceivability involved, see chapter 5. This notion of conceivability—conceivable on ideal rational reflection—also appears in Chalmers 1999a: 477n.

In being skeptical towards the evidential value of conceivability intuitions in philosophical arguments, Nagel and Loar partake in a well-established philosophical tradition. Mill (1974 [1843]: 238) writes:

> [T]here is such ample experience to show, that our capacity or incapacity of conceiving of a thing has very little to do with the possibility of the thing in itself; but is in truth very much an affair of accident, and depends on the past history and habits of our own minds.[14]

More recently, P.S. Churchland (1998) has argued that facts about what we can and cannot imagine, such as the fact that we can imagine zombies, are epistemological facts about us—facts about what we can and cannot understand given our present state of scientific knowledge. However, she argues, they are not, and do not entail, metaphysical facts "about the nature of the reality of the universe" (1998: 42).

In order to determine the value of conceivability intuitions, we must address some difficult questions. First, what should we take it to *mean* that something is conceivable? Besides the concept of conceivability, how shall we best interpret the other concepts in the conceivability thesis? It is to these questions I shall now turn.

[14] Mill's argument concerns *inconceivability*, but is, according to Casullo (1979: 216), general enough to cover conceivability as well.

CHAPTER 3

THE CONCEIVABILITY THESIS

In this chapter, I shall discuss the conceivability thesis, which says that *conceivability implies possibility*. This thesis has often been used in philosophical arguments. With the conceivability thesis as a premise, philosophers have attempted to prove that God exists, that certain forms of psychophysical dualism obtain, and more.[1] However, as we saw in the previous chapter, such attempts have often been met with extensive criticism: many philosophers believe that conceivability evidence for possibility is unreliable or even irrelevant.

The structure of the chapter is as follows. In section 3.1, I provide a brief account of the origins of the conceivability thesis, as dealt with by three classical philosophers who have endorsed it. In sections 3.2–3.4, I present preliminary considerations regarding three main questions that arise in connection with the conceivability thesis. In particular, I distinguish three ways in which you can save this thesis in the face of various counterexamples and counter-considerations (section 3.3), and I adopt an interpretation of the term 'possible' (section 3.4).

3.1 The conceivability thesis and its proponents

At least since Descartes (1596–1650), the conceivability thesis has often been employed in philosophical arguments. In Meditation 6 of the *Meditations on First Philosophy*, Descartes famously argued that he could have existed without his body, and that he therefore is distinct from his body. His argument proceeds from a general formulation of the conceivability thesis: "I know that everything which I clearly and distinctly understand is capable of being created by God so as to correspond exactly with my understanding of it" (CSM 2, Med 6: 54).

---

[1] The conceivability thesis appears in most formulations of the ontological proof of God's existence, which is originally due to Anselm of Canterbury (see Plantinga 1974, chapter 10, especially: 199). Moreover, Kripke (1980) and Chalmers (1996) have presented two recent arguments for psychophysical dualism that depend on the conceivability thesis (see section 4.4.1 and 4.4.2).

Elsewhere, Descartes asserts even more directly that "possible existence is contained in the concept or idea of everything that we clearly and distinctly understand" (CSM 2, Rep 1: 83). We shall devote a large part of chapter 4 to the study of Descartes' argument and the views concerning the conceivability thesis held by him and his critics.

Other philosophers have also entertained versions of the conceivability thesis. In his *Treatise of Human Nature*, Hume wrote:

> 'Tis an establish'd maxim in metaphysics, That *whatever the mind clearly conceives includes the idea of possible existence*, or in other words, *that nothing we imagine is absolutely impossible*. We can form the idea of a golden mountain, and from thence conclude that such a mountain may actually exist. We can form no idea of a mountain without a valley, and therefore regard it as impossible. (1978 [1739]: 32, Hume's italics)

The first italicized statement is similar to Descartes' claim that possible existence is contained in the concept or idea of everything that we clearly and distinctly understand. However, Hume's thesis is stronger than Descartes'. Whereas Hume and Descartes agree that *conceivability implies possibility*, Hume also makes the converse claim: that *possibility implies conceivability*. Therefore, the thesis "S is possible iff S is conceivable" has become known as *Hume's thesis* or *Hume's maxim* (see for example Sorensen 1992: 35; Häggqvist 1996: 126). And the early Wittgenstein employed a version of the conceivability thesis in the *Tractatus*.

> Der Gedanke enthält die Möglichkeit der Sachlage, die er denkt. Was denkbar ist, ist auch möglich. (1922, §3.02)

Since Descartes, Hume, and Wittgenstein entertained different concepts of conceivability and possibility, they also entertained different versions of the conceivability thesis. Both Descartes and Hume had *conceptual* notions of possibility,[2] but Descartes took conceptual possibility to coincide with what is possible in the world. On most interpretations, Hume did not accept that there is such a thing as natural, as opposed to logical or conceptual, necessity (Smith 1992: 301; De Pierris 2002: 499). In the beginning of his *Enquiries Concerning Human Understanding*, he says that for any philosophical term employed, we should ask ourselves the following question: *from what impression is the idea (represented by the term) derived?* (Hume 1975 [1777]: 22). When it comes to the idea of natural necessity, Hume says,

> we are never able, in a single instance, to discover any power or necessary connection; any quality, which binds the effect to the cause, and renders the one

[2] For this interpretation of Descartes, cf. Bennett 1994.

> an infallible consequence of the other. We only find, that the one does actually, in fact, follow the other. (Hume 1975 [1777]: 63)

The general point here is that in attributions of necessity to external states of affairs, the idea of something obtaining by necessity does not derive from any impression of these states of affairs. Rather, the idea of necessity derives from an expectation or "determination of the thought" following experiences that have certain features (cf. Stroud 1995: 466–7). For example, when we observe that two events A and B stand in "constant conjunction" with each other, we are inclined to infer that A *necessitates* B. In *Treatise of Human Nature*, Hume writes:

> Upon the whole, necessity is something that exists in the mind, not in objects; nor is it possible for us ever to form the most distant idea of it, consider'd as a quality in bodies. Either we have no idea of necessity, or necessity is nothing but that determination of the thought to pass from causes to effects and from effects to causes, according to their experienc'd union.[3]

Although Hume rejects the notion of natural necessity, he seems, in this passage, to acknowledge the existence of a *different* type of necessity, namely the type of necessity we have referred to as "broadly logical" or "conceptual" necessity. Hume's understanding of this type of necessity is often described as "psychologistic" (Pap 1958: 84). On Hume's view, necessary relations between ideas (or concepts) do not obtain independently of thought, but are contributed by the mind (cf. Hochberg 1995: 499). This view comes out explicitly in Hume's claim that

> the necessity, which makes two times two equal to four, or three angles of a triangle equal to two right ones, lies only in the act of the understanding, by which we consider and compare these ideas […].[4]

In the *Tractatus*, Wittgenstein agreed with Hume that there is no causal necessity (1922, §§5.135–5.1361), and he further argued that logical necessity is the only type of necessity there is (1922, §6.37).[5] Wittgenstein defended a form of *logical atomism* according to which the substance of the world is made up of simple things (§2.021), and argued that the logical form of each simple thing determines which states of affairs the thing can enter into (§2.014).[6] All

---

[3] Hume 1978 [1739]: 165–6. This passage is quoted in Stroud 1995: 466 and Hochberg 1995: 500.

[4] Hume 1978 [1739]: 166. The passage is quoted in Hochberg 1995: 505.

[5] See also Stenius 1996: 58–9, on the relation between Wittgenstein's and Hume's treatment of the notion of causal necessity.

[6] §2.0141 reads: "Die Möglichkeit seines Vorkommens in Sachverhalt ist die Form des Gegenstandes." Stenius argues that §2.0141 has the following meaning: "(a) that what

configurations of simple things (that is, states of affairs) are held to obtain independently of each other (§§5.134–5.135). Hence, Wittgenstein entertained what may be called a *combinatorial* concept of possibility: simple things can be conjoined into state of affairs, and those states of affairs obtain (or do not obtain) independently of each other. In an attempt to develop a theory of possibility that is Wittgensteinian in spirit, Skyrms (1989) writes:

> For possible worlds whose objects and relations are subsets of this world our possibilities are essentially *combinatorial*. We rearrange some or all of our relationships between some or all of the objects to get our possibilities. It is possible for the mat to be on the cat. (Of course not all combinations count as possible—the mat cannot be larger than itself—but I think that the reasoning involved here is basically combinatorial.) […] Wittgenstein believed that this is the *only* conception of possibility that we can have, and indeed that all possible worlds must contain exactly the same objects. (1989: 147–8)

As Skyrms implies, it is easy to see that in order for our notion of possibility to be "purely" combinatorial, all possible worlds must contain exactly the same objects. If some possible world contains more (or other) objects than there actually are, this cannot be accounted for in terms of re-combinations of the actual objects (cf. Skyrms 1989: 148).

In one sense, Wittgenstein can be understood as denying that things in the world have modal properties, and his views could in this respect be taken to resemble Hume's. By "things," we then mean ordinary things such as tables, chairs and people. However, it is sufficiently clear that Wittgenstein's *simple* things have modal properties. Firstly, the simple things are said to make up the *substance* of the world. In Black's (1964: 9) terms, the simple things "are what they are independently of what contingently is the case" (that is, independently of what facts that do or do not obtain). From this, we can conclude that the simple things exist by *necessity*: in Wittgenstein, a description of what facts that do or do not obtain is equivalent to a description of the *world*. Hence, given that the simple things are what they are irrespective of, e.g., what is the case in the world, simple things exist by necessity. Secondly, Wittgenstein argues that the (logical) *form* of a simple thing determines what states of affairs it *can* enter into. In other words, the logical form of a simple thing determines the (combinatorial) *possibilities* of that thing to enter into states of affairs. The logical form of a thing could perhaps also be called the *essence* of the thing (cf. Wittgenstein 1922, §§2.012; 2.0124, 2.0141, and Black 1964: 57).

---

decides which states of affairs a 'thing' can enter into is called its '(logical) form' and conversely that (b) a thing's logical form' is determined by the atomic states of affairs into which it *can* enter" (1996: 67–8).

The differences between Descartes, Hume, and Wittgenstein when it comes to their concepts of possibility point to a set of substantial problems concerning how to interpret the conceivability thesis. Not only do different philosophers entertain different notions of possibility, they also provide different interpretations of 'conceivable' and 'implies'. Chalmers notes that the term 'conceivability' can be understood in many ways (2002a: 146), and that on some interpretations of 'conceivable', an entailment relation from conceivability to possibility is "out of the question," and on other interpretations, "things are not so clear." There are also different possible interpretations of 'implies' in "conceivability implies possibility." 'A implies B' may be taken to mean that A constitutes *fallible evidence* for B, that A constitutes *sufficient evidence* for B, or that A *logically entails* B, etc.

In summary, there are three main problems that have to be addressed before we can come to grips with the question whether the conceivability thesis is true or even plausible. The three problems are:

(1) What should we take it to mean that something is conceivable?

(2) What should we take "implies" to mean?

(3) What kind of possibility shall we take the conceivability thesis to attribute to conceivable things?

It is obvious that our decisions about how to address these problems will have a decisive bearing on later attempts to address the conceivability thesis. In sections 3.2–3.4, I shall present preliminary considerations regarding (1)–(3).

3.2 What should we take it to mean that something is conceivable?

Among the alternative interpretations of 'conceivable' that Chalmers alludes to above, we also find a number of attitudes towards the question of what it means that something is conceivable. Some philosophers seem to think that there is no fact of the matter that determines what 'conceivable' means, or that the term 'conceivable' has no clear sense. If they are correct, it is impossible to provide a definition of what 'conceivable' means. However, one may still believe that it is possible to give an *explication* of the term à la Tarski, where one (i) indicates the purpose one wants the term to fulfill, (ii) provides adequacy conditions that any term must meet in order to fulfill this purpose, and (iii) attempts to give a definition that meets the adequacy conditions. With respect to the term 'conceivable', we shall see below what such adequacy conditions might be.

Other philosophers seem to think that this question has a determinate answer, although this answer is ambiguous between a set of established (philosophical and everyday) senses of 'conceivable'. For example, suppose that someone argues that a statement (or description) S is conceivable just in case S is *imaginable*, and that someone else argues that S is conceivable just in case S is *consistent* or *coherent*. Then, on this conception, they are both entitled to their interpretations, since "S is imaginable," as well as "S is consistent" and "S is coherent" are, to some extent, established senses of 'conceivable'.

A third group of philosophers seem to presuppose that you can provide a unique definition of 'conceivable'. This view, if correct, could undermine the conceivability thesis. Suppose for instance that a person believes that "is conceptually consistent" is the correct analysis of 'conceivable'. Since there are many scenarios that are strictly impossible (in a substantial sense to be outlined below) but nevertheless conceptually consistent, this person can reject the conceivability thesis as patently false.

In this chapter, I shall adopt a mixture of the attitudes referred to above. First, I do not believe that a unique analysis of what it means that something is conceivable is possible: there are many senses of 'conceivable', none of which is *the* right one. Secondly, I believe that if I were to adopt the position of the third group of philosophers, this would not allow an open-minded discussion regarding the alternative interpretations of 'conceivable' that appear in the philosophical literature. Thirdly, I cannot accept just any (unrestricted) interpretation of 'conceivable'. My primary interest is in the conceivability thesis, and this interest will restrict my discussion of different interpretations of 'conceivable'. Finally, I shall not require that an interpretation of 'conceivable' *must* be consistent with some of the established senses of 'conceivable'. Even a philosophical interpretation of what it means that something is conceivable must have some connection with what we normally might mean by "conceivable," but it need not be consistent with any particular interpretation. For example, consider Yablo's definition of what it means that something is conceivable (1993: 29):

(4) A statement S is conceivable for a conceiver *a* if *a* can imagine a world that *a* takes to verify S.[7]

---

[7] I have made some terminological changes in Yablo's definition: I speak of statements where Yablo speaks of propositions. (I shall often take the liberty to do so below.) One can say that Yablo's definition is formulated in terms of belief since "takes it that" can roughly be interpreted in terms of belief (a suggestion made to me by Yablo in personal communication).

Yablo's definition involves certain technical terms, and it is more elaborate than any everyday interpretation of what it means that something is conceivable. Nevertheless, it is formulated in terms one associates with many everyday interpretations: in terms of imagination, belief, and so on. In this sense, even a philosophical interpretation of 'conceivable' will have something in common with the everyday senses of the term.

3.3 What should we take "implies" to mean?

There are many ways in which one could interpret "implies" in the conceivability thesis. First, one could take the conceivability thesis to be a *global* truth, or one could take it to be a *local* truth. If one takes the conceivability thesis to be a global truth, one (roughly) takes the thesis to be true in any circumstances. What this amounts to depends on what strength one assigns to "implies." If one takes "*A* implies *B*" to mean "*A* is a fallible guide to *B*" (see below), then one take conceivability to be a fallible guide to possibility *in any circumstances*. On the other hand, if one takes the conceivability thesis to be a local truth, then one takes the conceivability thesis to be true only in *special circumstances*. For example, if one takes "*A* implies *B*" to mean "*A* is a fallible guide to *B*," one only takes conceivability to be a fallible guide to possibility in certain circumstances (as opposed to understanding it as an infallible guide to possibility, or as no guide at all, in *other* circumstances).

Secondly, one could take "implies" to mean that conceivability *suffices* for possibility, or one could take it to mean that conceivability provides *fallible evidence* for possibility. If one takes "implies" to mean that conceivability suffices for possibility, one takes it that if something is conceivable, then it is also possible. On the other hand, if one takes "implies" to mean that conceivability is fallible evidence for possibility, one takes it that modal judgments based on conceivability may sometimes be mistaken.

In this thesis, I shall tentatively accept a strong interpretation of the conceivability thesis. According to this interpretation,

(i) The conceivability thesis is a *global* truth

(ii) Conceivability is *sufficient* or *infallible* evidence for possibility.

One should at least investigate into the question whether such a strong interpretation is workable before considering other interpretations. Yablo (1993) does not think that the strong interpretation of "implies" in (i) and (ii) is reasonable. Instead, he argues that the conceivable is *ordinarily* possible, or that conceivability is *fallible evidence* for possibility. Other authors have attempted

to defend versions of the conceivability thesis similar to (i)–(ii). For example, Chalmers (2002a) holds that certain special forms of conceivability *entail* possibility (see section 5.3.3). What is distinctive of the types of conceivability Chalmers takes to *entail* possibility is that they involve what may be described as idealizations. For example, Chalmers takes conceivability based on *ideal rational reflection* to entail possibility.[8] Prima facie, this idea seems to involve certain idealizations.

Yablo and Chalmers thus employ different strategies in order to meet the counterexamples that have been presented against the conceivability thesis. Yablo's idea is the following: since we are imperfect reasoners, we will sometimes find impossible statements conceivable. Hence, it is unreasonable to expect *all* applications of the conceivability thesis to be valid. Thus, we can at best characterize conceivability as fallible evidence for possibility. However, in the absence of evidence to the contrary, such evidence provides us with prima facie entitlement to believe that what we have conceived of is possible. In summary, Yablo (roughly) denies (ii), but attempts to maintain (i).

Chalmers employs another strategy. He agrees that we are imperfect reasoners, and that we will sometimes find impossible statements conceivable. Therefore, there will be counterexamples to the conceivability thesis. However, Chalmers does not want to give up (ii), as Yablo does. Let us therefore roughly consider a reasoner or a reasoning process that is not subject to the cognitive restrictions that hamper the reasoning of a normal human conceiver. It is reasonable to suppose that such a reasoner or reasoning process will detect at least some of the impossibilities that we fail to detect. In the limiting case, a *perfect* reasoner should be able to detect all the impossibilities that we fail to detect. (Chalmers does not want to go as far as to consider a *perfect* reasoner.) In other words, Chalmers attempts to maintain both (i) and (ii), but this requires certain idealizations.

I shall proceed by considering another strategy one may adopt with regard to the question of how you should interpret "implies," in connection with possible counterexamples to the conceivability thesis. I shall further contrast this strategy with Yablo's. I believe that there is a difference between arguing, as Yablo does, that conceivability is a (fallible) *guide* to possibility, and arguing that the conceivability thesis is a *local truth*. The claim that conceivability is fallible evidence for possibility amounts to a denial of (ii), but it is nevertheless compatible with an acceptance of (i). On the other hand, the claim that the conceivability thesis is a local truth is compatible with an acceptance of (ii), but

---

[8] See Chalmers 1996: 68, 98, 100, and 131; Chalmers 1999b: 436; Chalmers 1999a: 477n.

amounts to a denial of (i). My arguments will bear on van Inwagen's paper "Modal Epistemology" (1998). In this paper, van Inwagen discusses Yablo's 1993 paper at length, but he does not seem to consider what I believe to be the case, namely that his arguments have another import than Yablo's.

In his paper, van Inwagen sees an analogy between modal judgments and judgments of distance made by sight. Many judgments of distance made by eye are approximately correct. Such judgments include "That chair is two feet from the wall," and "It is about one mile to that hill." However, van Inwagen notices, the practice of judging distance by eye is sometimes applied to cases in which it results in gross mistakes. For example, on a bright night it may appear as if the moon were just a couple of miles away:

> People had no idea how far away the sun and the moon and the stars were till they gave up trying to judge celestial distances by eye and began to reason. ("You can see a significant portion of the shadow of the whole earth on the moon when the moon is entering or leaving the earth's shadow, so the moon must be a lot farther away than anyone would have guessed …") (1998: 70)

The point of this example is the following. There are circumstances in which the practice of judging distances by sight yields correct results, and other circumstances in which it does not. These circumstances can also be described in general terms: the practice yields correct results when applied to objects in the vicinity, whereas it does not yield correct results when applied to very distant objects, such as celestial bodies. By analogy: there are circumstances in which the conceivability thesis yields correct results (that is, circumstances in which conceivability entail possibility), and there are circumstances in which it does not. van Inwagen seems to argue that these circumstances can also be described in general terms: the conceivability thesis yields correct results when applied to things which are familiar to us from everyday practice ("that chair could have been three feet from the wall"). When it comes to states of affairs that are not thus familiar, we should not attempt to apply the conceivability thesis.

I claim that van Inwagen's example can be taken to promote the thesis that conceivability implies possibility is a *local truth*, and that conceivability, in the right circumstances, suffices for possibility. To my mind, this is *not* the same thing as saying that conceivability is a *guide* to possibility. In terms of van Inwagen's example, the thesis that conceivability is a *guide* to possibility should rather be described as follows. Consider all the different types of distances we considered above: distances between objects in our vicinity, distances between celestial bodies, and so on. On the "guide" interpretation, we would say that the practice of judging distance by eyesight is a *guide* to the actual distances involved. Sometimes, we are correct (such as in the case with the chair and the

wall), and sometimes we are wrong (such as in the case with the moon), and when we are wrong, our judgments can be defeated by auxiliary evidence (for example, evidence obtained by geometrical reasoning).

In summary, Yablo proposes that we give up (ii) but maintain (i). (In other words, he sacrifices reliability for universalizability.) Chalmers attempts to defend both (i) and (ii), but this requires idealized notions of conceivability. Finally, van Inwagen maintains (ii) but rejects (i). (In other words, van Inwagen sacrifices universalizability for reliability.) Below, we shall see another version of van Inwagen's strategy when we discuss Kripke's mind-body argument in *Naming and Necessity*.

In summary, the possible interpretations of the conceivability thesis, based on the above distinctions, are:

(5) The conceivability thesis is a global truth, and conceivability is a sufficient (infallible) guide to possibility.

(6) The conceivability thesis is a global truth, but conceivability is a fallible guide to possibility.

(7) The conceivability thesis is just a local truth, but conceivability is a sufficient (infallible) guide to possibility.

(8) The conceivability thesis is just a local truth, and conceivability is a fallible guide to possibility.

I shall initially defend (5), but then proceed to point out certain difficulties pertaining to it. As regards (8), it is difficult to find any explicit proponents. This may simply be because (8) is so weak that it appears uninteresting.

3.4 What kind of possibility shall we take the conceivability thesis to attribute to conceivable things?

To begin with, I shall make two distinctions pertaining to the concept of possibility. First, one can distinguish between *objective* and *non-objective* (or *subjective*) notions of possibility. On objective notions of possibility, modal facts obtain independently of conceivers. For example, some philosophers believe that logical laws are features of an abstract but objective reality. Given such a concept of logic, the notion of *logical possibility*

(9) S is logically possible iff $S \nvdash \bot$ (in first-order logic).

is an objective notion of possibility. Likewise, on some theories of the nature of concepts, concepts are abstract, mind-independent entities, and *conceptual truth*

concerns the way in which concepts are related to each other. Given such a theory of concepts and conceptual truth, a notion of *conceptual possibility* according to which

(10) S is conceptually possible iff ¬S is not a conceptual truth

is an objective notion of possibility. Objective notions of possibility further include, at least on the standard conception of them, metaphysical and physical possibility.

On non-objective (or subjective) notions of possibility, on the other hand, modal facts are ultimately explained in terms of properties of agents. In our discussion of *modal eliminativism* in section 1.2.1, we saw Quine argue that things or states of affairs do not have modal properties independently of how we think of them or describe them. On this view, modal facts are subjective. Similarly, some philosophers hold that logical laws should be identified with or reduced to psychological laws, and that no other facts or laws need to be invoked to explain why *these* psychological laws, rather than any other set of laws, obtain. Given such a concept of logic, it seems as if, for example, the notion of logical possibility is subjective. Further subjective notions of modality are the notions of *epistemic* and *doxastic possibility*, which refer to consistency with what an agent knows or believes.

The second distinction I shall make is between *strict* and *non-strict* notions of possibility. There are many conceptions of what it means that something is possible. We encountered some of these in our introductory discussion concerning Descartes, Hume and Wittgenstein, and in section 2.3 above, where we discussed logical and conceptual analysis. Some of these notions are objective and some are not. Some of the objective notions are *strict* notions of possibility. As we noted in section 2.3, some philosophers think that there is an "unqualified" sense of 'possible', such that all other notions of possibility can, in some sense, be derived from this sense. It is with this sense of 'possible' in mind that we speak of "*strict* possibility," and it is probably with the same notion in mind that van Inwagen (1998) speaks of "possibility *simpliciter*" and Kripke (1980) of "possibility *tout court*" or "metaphysical possibility."[9] As I shall use the term, a notion N of possibility is a *strict* notion if the following holds:

---

[9] Kripke does not refer to his concept of possibility as "metaphysical" possibility, but as "possibility *tout court*" or just "possibility." However, as regards the concept of necessity, Kripke asserts that "what I am concerned with here [in *Naming and Necessity*] is a notion which is not a notion of epistemology but of metaphysics, in some (I hope) nonpejorative sense" (1980: 35–6).

(11) A statement S is possible with respect to N iff the world could have been such that S was true.

What "the world could have been such that S was true" precisely means (and what it is that determines how the world could have been) is a topic for chapter 6. Here, we shall rest content with saying that, for example, the sentence 'Water is not H_2O' involves no contradiction, and that it is therefore *logically possible* that this sentence could have been true. But could the world have been such that water would not be H_2O? Most philosophers agree that this could not have been the case. Thus, *logical possibility* is not a strict notion of possibility. Furthermore, it is often argued that it is conceptually coherent to suppose that water could have been something else than H_2O.[10] Thus, the demand for conceptual consistency is too weak if we want a strict notion of possibility: conceptual coherence does not always entail strict possibility.

3.4.1 Possibility as strict possibility

In this thesis, I shall take the conceivability thesis to attribute strict possibility to conceivable things. In other words, I shall take the conceivability thesis to be equivalent to the following claim:

(12) If S is conceivable, then S is strictly possible.

The main reasons for using the term 'strict possibility' instead of the more common Kripkean 'metaphysical possibility' are (i) to dissociate the notion of strict possibility from well-known Kripke-inspired views about what *makes* certain statements metaphysically possible, and (ii) because in chapter 6, I intend to use the term 'metaphysical possibility' in a more detailed sense. There is no absolute consensus among philosophers about what it means that something is metaphysically possible. What is clear is that whether something is metaphysically possible or not depends on the nature of the things in the world, and (in some cases), on the nature and history of the whole world. Scientific identifications, such as the identification of water with H_2O, the identification of gold with the atomic number 79, and so on, are all metaphysically necessary. But there are also metaphysical necessities that pertain to the way in which individual things have come into being. For example, the table at which I sit originates (in part) from a specific plank of wood. Since the table would have been another table had it originated from

[10] At least, it was conceptually coherent to make this supposition before we learned that water *is* H_2O. One might suspect that "being H_2O" belongs to the *contemporary* concept of water, but I do not think that it does.

another plank of wood, it is a metaphysically necessary fact about this table that it originates from the plank of wood it actually originates from.

In so far as Kripke's account of metaphysical possibility—which describes possibility as truth in at least one possible world—can be separated from the Aristotelian essentialism that many interpreters have attributed to him, then "strict" and "metaphysical" possibility are the same thing. However, the notion of strict possibility is explicitly intended to be compatible with different ways of explaining what "the world could have been such that S was true" means, that is, what we take it to mean for a world to be possible.

On the interpretation of "the world could have been such that S was true" that we shall provide in chapter 6, the notion of strict possibility introduced in (11) is, as we shall say, *autonomous*. We shall say that a concept of possibility is *autonomous* if what it means that something is possible is, relative to that concept, in no part defined in terms of conceivability. One interpretation of what it means for something to be possible which is not autonomous in the sense intended can be found in Lewis and Langford's *Symbolic Logic* (1959 [1932]: 160–61). Lewis and Langford are concerned with how to explicate the meaning of the expressions '$\Diamond p$', '$\neg\Diamond p$' and '$\neg\Diamond\neg p$', which are equivalent to "it is possible that *p*," "it is not possible that *p*," and "it is necessary that *p*" in the notation of symbolic modal logic. One of Lewis and Langford's suggestions about how to explicate the meaning of the possibility symbol '$\Diamond$' is the following: '$\Diamond p$', that is, "it is possible that *p*," should be taken to mean that *p* is *logically conceivable*. Thus, in one sense, Lewis and Langford define possibility in terms of conceivability. Their approach can be contrasted with the concept of metaphysical possibility. For something to be metaphysically possible, it must be compatible with objective and mind-independent facts about the world itself. On this view, conceivability does *not* determine possibility.[11]

[11] Note, however, that this is not to deny that there could be *some* relation between conceivability and possibility. Suppose that I am observing a man in the street from my window. Clearly, the fact that I observe this man is not what *makes* him stand there—he could have been standing there without anyone observing him standing there. On the other hand, the fact that I observe him standing there is evidence for the claim that he *is* standing there. The relation between conceivability and possibility might be the same: the fact that I can conceive of something is not what *makes* it possible. On the other hand, the fact that I find it conceivable might be taken as evidence for the claim that it is possible.

3.4.2 Why strict possibility?

Why should we interpret the conceivability thesis in terms of (12)? That is, why should we take the conceivability thesis to attribute strict possibility to conceivable things? The reason is that (12) is the version of the conceivability thesis that is employed in most philosophical arguments. In the philosophical arguments which make use of the conceivability thesis, the intention is usually to establish that something could have been the case with respect to the world itself, not merely to establish that something is possible in some formal sense. In order to substantiate this claim, I shall in the next chapter turn to three arguments in the philosophy of mind which employ the conceivability thesis.

CHAPTER 4

CONCEIVABILITY ARGUMENTS IN THE PHILOSOPHY OF MIND

4.1 Introduction

In the previous chapter, I argued that we should take the conceivability thesis to ascribe strict possibility to conceivable things. I further argued that a statement S is strictly possible if the world could have been such that S was true. In other words, I shall take the conceivability thesis to be equivalent to the following claim: *if* S *is conceivable, then* S *is strictly possible*. I also made the preliminary decision to take the conceivability thesis to assert that conceivability is a *global* and *sufficient* guide to possibility. During the course of the remaining chapters, we shall see to what extent this strong interpretation can be defended.

In this chapter, I shall indicate how conceivability considerations enter into arguments in the philosophy of mind. The purpose of the discussion to follow is to substantiate the claim that the above interpretation of the conceivability thesis is the interpretation that is needed in order to support most arguments in which inferences from conceivability to possibility occur. I shall discuss Descartes' mind-body argument in the *Meditations* (section 4.2), the mind-body argument Kripke presents in *Naming and Necessity* (section 4.4.1), and Chalmers' Zombie argument (section 4.4.2). The focus will be on Descartes' mind-body argument. I shall evaluate the arguments presented by Descartes and Arnauld in their exchange pertaining to this argument, and I shall also consider Yablo's (1990) assessment of Descartes' argument and his reply to Arnauld's criticism.

Another purpose of addressing Descartes, Chalmers, and Kripke is to map out the problems that arise in connection to claims about conceivability. In chapter 3 I presented some initial considerations regarding the problems of interpreting the conceivability thesis. In the present chapter, I shall attempt to make our conception of these problems more complex via the discussion of Descartes', Chalmers' and Kripke's arguments. In particular, I shall introduce and discuss the distinction between *genuine* and *apparent* conceivability which, explicitly or implicitly, appears in most discussions regarding the concept of conceivability and the conceivability thesis.

4.2 Descartes' mind-body argument and Arnauld's criticism

Descartes' main argument for mind-body dualism appears in Meditation 6 of the *Meditations on first philosophy* (CSM 2, Med 6: 54), and is sometimes referred to as Descartes' *separability argument* for mind-body dualism.[1] Descartes says the following:[2]

> First, I know that everything which I clearly and distinctly understand is capable of being created by God so as to correspond exactly with my understanding of it. Hence the fact that I can clearly and distinctly understand one thing apart from another is enough to make me certain that the two things are distinct, since they are capable of being separated, at least by God. […] Thus, simply by knowing that I exist and seeing at the same time that absolutely nothing else belongs to my nature or essence except the fact that I am a thinking thing, I can infer correctly that my essence consists solely in the fact that I am a thinking thing. It is true that I […] have a body that is very closely joined to me. But nevertheless, on the one hand I have a clear and distinct idea of myself, in so far as I am simply a thinking, non-extended thing; and on the other hand I have a distinct idea of a body, in so far as this is simply an extended, non-thinking thing. And accordingly, it is certain that I am really distinct from my body, and can exist without it.

In what follows, I shall attempt to explicate what I believe to be Descartes' main line of thought, and formulate an interpretation of his argument that attempts to account for these passages. Contemporary explications of these passages sometimes differ a lot as regards what they take its argument to be. In the following, I shall let the context determine whether "I" refers to Descartes or to the author of this text. I shall also take certain liberties in replacing various terms with others, and in adding clarifying comments. The reader should compare my premises with Descartes' own words.

Descartes' argument begins with a general formulation of the conceivability thesis. He asserts:

(A) What I clearly and distinctly understand is capable of being created by God so as to correspond exactly with my understanding of it.

(A) is usually taken to be equivalent to:

(1) What I can conceive clearly and distinctly is possible.

---

[1] Descartes is sometimes associated with other arguments for dualism. Wilson (1978: 190) argues that some of them can properly be ascribed to Descartes, such as "the argument from divisibility" (see Bennett 2001a: 69), but that others, such as the "argument from doubt," cannot be ascribed to Descartes.

[2] The arrangement of the relevant passages is adopted from Bennett (2001a: 71).

Descartes justifies (1) in two different ways. First, we are given a pragmatic justification: if (1) is rejected, modal knowledge cannot be attained (CSMK: 203). Secondly, we are given a theological justification: if (1) is false, then God is a deceiver (CSMK: 203).

In the next sentence, Descartes focuses on a special instance of (1) that will be important in his argument:

(B) the fact that I can clearly and distinctly understand one thing apart from another is enough to make me certain that the two things are distinct, since they are capable of being separated, at least by God.

We take (B) to be equivalent to:

(2) If I can conceive clearly and distinctly of x as being distinct from y, then x is distinct from y.

(2) requires some explanation. In the latter part of (B), the idea is that if x and y are such that God is capable of separating x from y, that is, if x and y are *possibly distinct*, then they are *actually* distinct. Thus, (2) follows from (1) via the auxiliary premise:

(*) $\forall x \forall y(\Diamond(x \neq y) \rightarrow x \neq y)$.

Those familiar with Kripke's thesis that identity is a necessary (or "internal") relation between objects (see Kripke 1980: 3), that is,

$\forall x \forall y(x = y \rightarrow \Box(x = y))$ (*Necessity of identity*)

can see that (*) follows from this thesis.

The next steps in the argument involve the notion of *essence*. Descartes seems to hold that our conceptions of certain things are such that the *essences* of the things in question are revealed to us in the conceptions themselves or by means of conceptual analysis. Veridical "essence reflection" of this kind can only occur if, and when, the conception (of the thing in question) is *clear and distinct*. Alanen (1996: 11) argues that for Descartes, *clearness* "is a question of presence, of how well or vividly or attentively the idea is perceived or entertained." *Distinctness* is, like clearness, a matter of degree, and is defined partly in terms of the concept of clearness: "the more attributes or properties of a thing we perceive clearly, the more distinct is our idea of it" (Alanen 1996: 11). We note that this also seems to impose a restriction on the *thing* that is the object of conception: it must not be so complex that we *cannot* entertain a clear and distinct conception of it, due to deficiencies in our conceptual and cognitive powers.

Descartes clearly holds that the *self*, or the "I" in the *Meditations*, is one of the things of which we can have the aforementioned kind of conception. The *body* is another such thing.[3] About the self and the body, Descartes asserts:

(C) I […] have a body that is very closely joined to me. But nevertheless, on the one hand I have a clear and distinct idea of myself, in so far as I am simply a thinking, *non-extended* thing.

Furthermore,

(D) I have a distinct idea of a body, in so far as this is simply an extended, *non-thinking* thing.

How does Descartes arrive at (C) and (D)? The support for (C) is the following:

(E) simply by […] seeing […] that absolutely nothing else belongs to my nature or essence except the fact that I am a thinking thing, I can infer correctly that my essence consists solely in the fact that I am a thinking thing.

In other words, by means of essence reflection:

(3) I clearly and distinctly understand that *nothing else* belongs to my essence except that I am a thinking thing.

Hence, by the veridicality of essence reflection:

(4) My essence consists *solely* in the fact that I am a thinking thing.

The idea behind (C) is that Descartes can entertain an idea of himself as a thinking and *non-extended* thing because it has been established by (4) that *nothing else* belongs to his essence *besides* being a thinking thing. If we generalize:

(**) For any property *F* (*being extended*; *being non-extended*) such that the possession of *F* is not incompatible with *being a thinking thing*, I can entertain the idea of myself as having the property *F*.

Now, we need the corresponding claims about the body. Unfortunately, there is no passage mentioning the essence of body that corresponds to (E) on which we can base claims about body that would correspond to (3) and (4). Nevertheless, Descartes must have entertained corresponding thoughts about body, because (D) (about the body) is perfectly analogous to (C) (about the self).

---

[3] In CSM 2, the editors note that the term 'corpus' used by Descartes is ambiguous between "body" as in corporeal matter in general, and "the body" as in *a* body, or *my* body (cf. CSM 2, Med 6: 54n).

Support for (D) would thus be the (tacit) claims analogous to (3) and (4), which are established by means of essence reflection regarding body:

(3*) I clearly and distinctly understand that *nothing else* belongs to the essence of (my) body except that it is an extended thing.

Hence, by the veridicality of essence reflection:

(4*) The essence of (my) body consists *solely* in the fact that it is an extended thing.

Now, from (C) and (D), one might be tempted to conclude via (2) that my body and I are distinct. However, if we take (C) to assert that I am conceivably *thinking and non-extended*, and if we take (D) to assert that my body is conceivably *non-thinking* and *extended*, this does *not* entail that it is jointly conceivable that I am *thinking and non-extended* and my body is *non-thinking* and *extended*. In other words, we take the following inference to be invalid:

$$\frac{\text{Conceivable}(S) \wedge \text{Conceivable}(S^*)}{\text{Conceivable}(S \wedge S^*).}$$

For example, it is (or so I shall suppose) conceivable that I did not exist, and it is also conceivable that I was seven feet tall, but it is not conceivable that I did not exist and still was seven feet tall.

Rather, we shall take Descartes separability argument to proceed from (4) and (4*), which are the premises on which (C) and (D) rely. Given that I know that my essence consists *solely* in the fact that I am a thinking thing, and that the essence of my body consists *solely* in the fact that it is an extended thing, I can *conceive* of a situation in which (i) I have none of the attributes that are essential to the body, and (ii) my body have none of the attributes that are essential to me. (The claim that I can conceive of such a situation is supported by (C) and (D), which state that I have an idea of myself as *non-extended*, and of my body as *non-thinking*.) However, to conceive of such a situation is to conceive of a situation in which I am distinct from my body, because in that situation, my body and I have incompatible essential attributes. (That is, if x and y have incompatible essential attributes, then x and y must have different essences. But if x and y have different essences, then x and y must be distinct things.) At this point, Descartes has established that he is conceivably distinct from his body. In other words:

(5) I can clearly and distinctly conceive of myself as being distinct from my body.

But then it follows from (2) that:

(6) I am distinct from my body.

Hence, we have reached the desired conclusion.

I shall indicate two additional arguments that build on the same passage in Descartes as the separability argument. What is particular to the following argument, which we shall call *the argument from essence*, is that it is compatible with positions that the separability argument is incompatible with. Consider the following argument: assume (3) and (3*). Thus, we get (4) and (4*). By (4), (4*) and essence reflection it follows that the property of being non-extended is compatible with *my* essence, but it is not compatible with the essence of *my body* (since the essence of my body consists *solely* in the fact that it is an extended thing). Hence, by reflection on what it means for (the possession of) a property *F* to be compatible with the essence of *x* but not with the essence of *y*, (6) follows. That is, if *F* is compatible with the essence of *x* but not with the essence of *y*, then *x* and *y* have different essences. But if *x* and *y* have different essences, then *x* and *y* must be distinct things.

One should note that the argument from essence does not rely on the separability premise (2). In fact, the argument from essence does not rely on considerations of conceivability and possibility at all, and it does *not* state that it is *possible* for me to exist without my body. Hence, this formulation of the argument seems to be compatible with recent theories that attempt to separate the notion of essence from the standard modal notions of possibility and necessity.[4] In this sense, the argument from essence is compatible with the claim that I am *necessarily* extended without being *essentially* extended. The idea is that I have a body in all possible worlds, but am nevertheless numerically distinct from my body, in the sense that I am a particular of a type of thing (a thinking thing), and my body is a particular of another type of thing (an extended thing).[5]

A third argument that builds on the same passage in Descartes as the separability argument offers an explication solely in terms of the general conceivability thesis (1). We may call the following argument *the conceivability argument*. (Like the argument from essence, the conceivability argument makes

---

[4] See chapter 6 below. Such theories are proposed by Kit Fine (1994) and Joseph Almog (1991).

[5] Such a theory is developed by Almog (2002a).

no use of the separability premise (2).) The argument starts out from (5) above. Here, (5) is a premise of the argument, whereas in the separability argument it is an intermediate step. From (1) and (5) it follows that:

(i) It is possible for me to be distinct from my body.

Furthermore,

(ii) It is not possible for my body to be distinct from my body.

In support of (ii), we take it to be a *brute logical fact* that any statement of the form $\neg\Diamond(a \neq a)$ is true. (This follows from the necessity of identity.) We thus take it to be a *brute logical fact* that *Descartes' body* cannot possibly be distinct from *Descartes' body*, and we do not need to enter into conceivability considerations in order to support (ii). Then, by (i) and (ii), it follows that:

(iii) I have a property that my body does not have, namely, the property of *possibly being distinct from my body*.

By Leibniz' law, it then follows that

(iv) I am distinct from my body.

That is: let a = Descartes and b = Descartes' body. Then *being possibly distinct from my body* is a property of a but not of b. But then the assumption that $a = b$ together with Leibniz' law yields a contradiction. Hence, $a \neq b$.[6]

I believe that neither of the last two explications we have considered amounts to the most charitable interpretation of Descartes. In the following, I shall therefore take the separability argument above to be the best approximation to Descartes mind-body argument.

4.2.1 Arnauld's criticism

In his objections to Descartes' mind-body argument (CSM 2, Obj 4), Arnauld raised what Descartes himself took to be the most important objection to this mind-body argument. In his objections, Arnauld questioned whether Descartes really did conceive of himself as disembodied. Arnauld argued as follows: Suppose that x claims to have conceived that Fy, and that this person therefore, via the conceivability thesis, concludes that it is possible that Fy. At this point (if discontented), the Arnauldian skeptic will raise the question whether this person *genuinely* conceived of Fy. Given that there is some reason to doubt this, the Arnauldian skeptic could *agree* with the conceivability thesis, but still *reject*

---

[6] The latter argument is inspired by (but not identical to) Almog's "argument from possibility." See Almog 2002a: 6–7.

the conclusion that it is possible that *Fy* since it has been questioned whether *Fy* was *genuinely* conceived.

Arnauldian skepticism thrives on a distinction between *genuine* and *apparent* conceivability. This distinction can be made in various ways. For example, we could say that for an conceiver *x* and a thing *y*, *x*'s conception of *y* is *genuine* just in case there is a suitable *causal* connection between *y* and *x*'s conception of *y*.[7] We shall also be concerned with different senses of 'apparently conceivable'. On the central notion of apparent conceivability that we shall be concerned with below, a statement S is apparently conceivable for a conceiver iff it appears to the conceiver that S is conceivable to him or her. Here, however, we shall be concerned with a sense of 'apparently conceivable' that we shall call *apparently conceivable in the Arnauldian sense*. In order for apparent conceivability in the Arnauldian sense to obtain, it must not only appear to the conceiver that something is conceivable for him- or her, the conceiver must also believe that he or she knows all the essential properties of the things involved.

In order for a particular conception to be *genuine*, Arnauld demands something else than a mere causal connection between the conceiver and the thing that is conceived. Notably, his requirement concerns the content of conception. Arnauld's requirement seems to be the following.[8] In order for it to be genuinely conceivable for *x* that, say, *Fy*, *x* must have an *adequate* conception or representation of *y*. In order for *x* to have an adequate representation of *y*, *x* must, for each essential property *G* of *y*, know that *G* is an essential property of *y*. If this criterion is fulfilled, we say that *x* possesses *adequate knowledge* of *y*. The general idea is that if *x* knows, for each essential property *G* of *y*, that *G* is an essential property of *y*, *x* will not attribute any property to *y* that is incompatible with *y*'s essential properties.

Whether *x*'s representation of *y* is adequate or not may be epistemically inaccessible to *x*. In other words, *x* may not know whether he or she in fact knows all of *y*'s essential properties. This is to say that the following does not always hold:

(7) $\text{KNOWS}_x[\forall G(y \text{ is essentially } G \rightarrow \text{KNOWS}_x(y \text{ is essentially } G))]$.

Suppose that for all *x* knows, it *might* be that *x* knows each of *y*'s essential properties, but it might also be that *x* does not. Suppose though, that *x believes*

---

[7] Compare (for example) Kaplan 1999: 381–2. Alternatively, we could say that genuine conceiving requires *acquaintance* with that which is purportedly conceived.

[8] The following interpretation relies mainly on Yablo 1990; 1993.

himself to have an adequate representation of *y*, and therefore a genuine conception of *y* that *Fy*. That is, suppose that the following obtains:

(8) $\text{BELIEVES}_x[\forall G(y \text{ is essentially } G \rightarrow \text{KNOWS}_x(y \text{ is essentially } G))]$.

If (8) is the case, *x* has a conception of *y* that is, as we shall say, *apparent in the Arnauldian sense*. In the following paragraphs, I shall use a star ('*') to indicate that I am concerned with the Arnauldian sense of 'apparently conceivable' as opposed to the sense of 'S is apparently conceivable to *x*' on which it is only required that it appears to *x* that S is conceivable to him or her.

Now, if *x* has an apparent* conception of *y*, *x*'s conception of *y* that *Fy* will be one of the following: either *x*'s conception of *y* that *Fy* is genuine (which entails that *x*'s belief represented in (8) is correct), or *x*'s actual conception of *y* that *Fy* is *merely apparent*, that is, misrepresentative and erroneous.

Arnauld's criticism of Descartes turns on the mere plausibility of the suggestion that Descartes' conception of himself that he could have existed without a body *might be* a merely apparent conception. Note that Arnauld only suggests that Descartes' conception *might be* a merely apparent conception—not that it *is* a merely apparent conception. Given that one does not know whether a particular conception is genuine or not, one cannot know that it *is* merely apparent either.

One should note that Arnauld never denies the conceivability thesis. What he argues is that the conception must be of the right kind. According to Arnauld, the legitimate version of the conceivability thesis is the following:

(9) *Genuine* conceivability implies possibility.

By contrast,

(10) *Apparent** conceivability does not always imply possibility.

Arnauld thus grants Descartes *apparent** conceivability, but the inference to possibility is still blocked, since the apparent* conceivability may be *merely* apparent conceivability. Arnauld's criticism thus comes down to the following. If Descartes apparently* conceives of himself in a disembodied state, this could very well mean that he could exist in such a state, but it might as well mean nothing, that is, Descartes' conceivability intuitions might have no modal import at all.

4.2.2 Descartes' reply to Arnauld

In his reply to Arnauld's objection, Descartes argues that the type of knowledge that Arnauld requires is not needed in order to make conceivability suffice for possibility:

> I do not, as M. Arnauld assumes, think that adequate knowledge of a thing is required here. […] A created intellect […] though perhaps it may in fact possess adequate knowledge of many things, can never know it has such knowledge unless God grants it a special revelation of the fact. […] Now in order for us to recognize a real distinction between two things it cannot be required that our knowledge of them be adequate if it is impossible for us to know that it is adequate. (CSM 2, Rep 4: 155)

In other words, Descartes holds that (7) cannot be a necessary requirement on modal knowledge claims.[9] However, he argues,

> […] it may be that there is much within me of which I am not yet aware […]. Yet since that of which I am aware is sufficient to enable me to subsist with it and it alone, I am certain that I could have been created by God without having these other attributes of which I am unaware, and hence that these other attributes do not belong to the essence of the mind. (CSM 2, Rep 4: 154–5)

In the terminology of the *Replies*, Descartes argues that for conceivability to be sufficient for possibility, all that is needed is that our knowledge of the thing to which we apply the conceivability thesis is *complete*. A complete understanding of something is acquired if one can understand a thing *y* to be a self-subsistent thing qua possessor of exactly the properties *M* (where *M* is a specific set of properties). Here it is important to note that Descartes takes the statement "understanding something to be a complete thing" (that appears to pertain solely to the nature of the *content* of the relevant conception) and the statement "a complete understanding of something" (that seems to pertain also to the nature of the *thing conceived*) to have the same meaning. Thus, Descartes seems to hold that if a thing *y* can be clearly and distinctly conceived as a complete thing

---

[9] In passing, we should note that at this point in the exchange between Descartes and Arnauld, they seem to misunderstand each other in one respect. Descartes seems to mean that *x* has adequate knowledge of *y* iff *x*'s knowledge of *y* includes absolutely all of *y*'s properties. By contrast, as we saw above, Arnauld seems to mean that *x* has adequate knowledge of *y* iff *x*'s knowledge of *y* includes all of *y*'s *essential* properties. As mentioned above, our interpretation of Arnauld relies on Yablo (1993). Yablo argues that Arnauld's concept of adequacy is the one mentioned in (7) above. However, Yablo is ambiguous between interpreting Arnauld as requiring knowledge of the essential properties of a thing *x*, and as requiring knowledge of *all* properties of *x* (see Yablo 1990: 159). If we were to adopt the latter interpretation, the difference between Descartes' and Arnauld's concepts of adequacy would disappear.

via the set of properties *M*, then *x is*, qua possessor of exactly the properties *M*, a complete thing (cf. Yablo 1990: 168–9).

According to Cartesian metaphysics, for *x* to exist with the *M* properties to the exclusion of all other properties, that is, for *x* to be a complete thing, is for *x* to be a *substance* (characterized by the *M* properties). To be a substance is to be capable of independent existence (cf. Yablo 1990: 164).

We now ask what is required in order for a thing *x* to be conceivable as (or, equivalently, to *be*) a complete thing qua possessor of a certain set of properties *M*. According to Yablo, Descartes holds that in order for a conception of *x* to be complete (or, equivalently, for *x* to be conceivable as a complete thing), *x* must be clearly and distinctly conceivable as possessing the properties *M* to the (positive) exclusion of all other properties (cf. Yablo 1990: 168). This requirement is meant to prevent mistaken modal attributions resulting from *abstraction*. To conceive of *x* that, say, *Hx*, *by abstraction* normally means leaving certain properties of *x* out of consideration in a way that makes the thought "*Hx*" coherent. For example, for all I know, it may be that I am essentially embodied. Nevertheless, I can leave any thoughts about my body out of consideration in my conception of myself and in this way enable a conception of myself as disembodied. Thus, by abstraction, it may happen that I subtract *essential* properties from that which is conceived, and thereby render the resulting conception *inadequate*. In his response to Arnauld, Descartes stresses that although conceptions need not be *adequate*, they must be such that they have *not* been rendered *inadequate* "by an abstraction of the intellect" (cf. CSM 2, Rep 4: 155; Yablo 1990: 165). Thus, in order for a conception to be complete and not rendered inadequate (in the sense described), *x* must be conceivable as having exactly the *M* properties, and, *x* must be conceivable as (positively) *lacking* all other properties.

Given these clarifications, I believe that the intended interpretation of the separability argument is the following. (This interpretation roughly follows Yablo 1990: 170.) Descartes' initial claim is that he has a complete conception of himself as possessing exactly what Yablo calls his *thought properties* (Yablo 1990: 151, 169; CSM 2, Rep 4: 156–7). One's *thought properties* are exactly the properties that one is immediately aware of possessing (Yablo 1990: 151). For example, I am immediately aware of my deliberations. Thus, *being a thinking thing* is one of my thought properties. By contrast, my awareness of my body is mediated by my senses. Thus, *being an embodied thing* is not one of my thought properties. From the conceivability thesis it thus follows that Descartes *can* exist (that is, *it is possible* for Descartes to exist) with exactly his

thought properties (cf. Yablo 1990: 170). However, if it is possible for a thing x to exist with exactly the properties M, then x is a substance (qua possessor of the M properties), and is not ontologically dependent on the existence of any other thing. In Descartes' case, this means the he, *qua* possessor of his thought properties, is a (thinking) substance and can exist without his body (cf. Yablo 1990: 170). Thus, from Leibniz' law, it follows that Descartes is distinct from his body.

To the contemporary modal epistemologist, this argument may seem no more convincing than the version of the argument that Arnauld called into question. Again, how can we be sure that modal judgments based on complete conceptions are veridical? Is there not room for Arnauldian doubt also in this case? However, Yablo argues that we can resist *gratuitous* attributions of error, and take conceivability to be prima facie evidence for possibility. At this point, Yablo attempts to shift the burden of proof back to the Arnauldian skeptic. Gratuitous attributions of error will not be accepted: the Arnauldian skeptic must point specifically to where he thinks that the modal reasoning has gone wrong. In order to explain what type of specific criticism he has in mind, Yablo cites the familiar example of Phosphorus and Hesperus (1990: 181–2). The ancients could conceive that Phosphorus was distinct from Hesperus, the story goes, since they did not know that Phosphorus = Hesperus = Venus. Here it is evident where the modal reasoning of the ancients has gone wrong: they did not know that Phosphorus = Hesperus. According to Yablo, the ancients' mistake can be explained in terms of the following *model of modal error* (cf. Yablo 1990: 182–5):

(i) q

(ii) $q \rightarrow \Box\neg p$, and

(iii) my ability to conceive it as possible that p is explained by my ignorance that (i), or else my ignorance that (ii).

Applied to the Phosphorus-Hesperus case, q is the proposition that Phosphorus is identical to Hesperus, and p is the proposition that Phosphorus and Hesperus are distinct. The ancients' apparent ability to conceive that Phosphorus is distinct from Hesperus is thus due to their ignorance of (i). Regarding the conclusion of the Cartesian separability argument, Yablo argues that

> the objector's challenge is to identify a proposition q for which there are *independent* grounds to suspect that my conceivability as a purely thinking thing is explained by my ignorance of the following fact: that q is a truth which shows that this is impossible for me. (1990: 188)

Thus, with respect to my modal intuition that I could exist in a disembodied state (which is supported by my complete conception of myself in such a state), the question is whether there is any proposition *q* such that the following holds:

(i) *q*

(ii) $q \rightarrow \Box$(I possess more properties than just my thought properties *F*, …, *G*), and

(iii) my ability to conceive it as possible that I should possess no more than my thought properties is explained by my ignorance of (i), or else by my ignorance of (ii).

The obvious candidate for *q* here, Yablo argues, is the proposition that I am *incapable* of purely mental existence. However, even if there is a significant chance that this proposition is true, or that another related proposition (the truth of which would make it impossible for me to exist with only my thought properties) is true, this could hardly count as an *objection*, Yablo argues:

> After all, it could equally be said that I am able to conceive it as possible that I should have had a different birthday, only because of ignorance about the necessity of my actual birthday. In either case, the most that can be claimed is that *if* the alleged defeater is true, and, e.g., it *is* necessary that I am born on September 30, then if I had not been ignorant of that fact, I would not have found any earlier birthday conceivable. And that is hardly an *objection*; no more than it is an objection to the veridicality of my perceptual impression that there are ducks present, that if I am wrong, and they are decoys, then my ignorance of that fact would figure in the explanation of how I was able to suppose that they were ducks. (1990: 187)

If we instead turn to possible explanations of *why* I am incapable of purely mental existence, we might mention the proposition that I am necessarily or essentially extended (Yablo 1990: 190), or *the identity thesis* which, when applied to me, in its crudest formulation, states that I am identical to my body or my brain (1990: 190). Both of these propositions entail that I am incapable of purely mental existence. Nevertheless, both of these propositions are themselves in want of independent evidence. The best we can do at the present state of scientific knowledge is to say that *if* the identity thesis is true, or *if* I am necessarily or essentially extended, *then* I am incapable of purely mental existence. And, again, on Yablo's view, this hardly counts as an objection.

It is interesting to note that the criticism of dualistic or otherwise anti-materialist arguments is often such that it could be described as *gratuitous* in Yablo's sense. For example, in his remarks regarding the zombie argument (see section 2.8), Nagel argues that the apparent conceivability of zombies is due to "limitations of our own understanding" (1998: 342). He further argues that we

"do not at the present possess the conceptual equipment to understand how subjective and physical features could both be essential aspects of a single entity or process" (1998: 342). Nagel implies that a conceptual revolution, with respect to the relevant concepts would provide us with equipment that would facilitate such an understanding.[10] Similarly, Patricia Churchland (1998) believes that it will eventually be possible to explain the mental in physical terms, and she implies that this belief is supported by general considerations concerning the development of science. However, neither Nagel nor Churchland presents what Yablo would count as an *objection* to the relevant conceivability intuitions. For example, Nagel implies that a single entity (e.g. the mind) can have both subjective and physical properties among its essential features, but that we, at the present, do not "possess the conceptual equipment to understand" how this is possible. This is precisely the type of general skepticism that Yablo refuses to count as an *objection* to the relevant conceivability intuition.

So where do we currently stand with regard to the mind-body problem in general, and Descartes' argument in particular? On the one hand, we have the amended separability argument, which attempts to meet Arnauld's criticism. According to the amended separability argument, conceptions need not be *adequate*, in the sense of (7), in order for conceivability to imply possibility—they only have to be *complete*. The corresponding interpretation of the conceivability thesis would thus be the following:

(11) Conceivability based on complete ideas entails possibility.

Furthermore, we have Yablo's attempt to shift the burden of proof back to those who are skeptical about (11).

On the other hand (pace Yablo), there is the worry that there is room for Arnauldian doubt also in this case. According to my understanding, Descartes idea is that *if* I have a complete conception of a thing *a*, *then*, even if I do not know all the essential properties of *a*, I would not (be able to) ascribe the property of being possibly *H* to *a* if *a* has an essential property which prevents *a* from being *H*. Descartes holds that a conception is complete (in the above sense) if it is clear and distinct and if it has not been rendered inadequate by an abstraction of the intellect. Both Descartes and Arnauld hold that one cannot know whether one's conception of a particular thing is adequate. However, it seems as if Descartes thinks that if a conception appears to be complete, then it *is* complete. At least, this is one way of interpreting his remark that he takes

[10] For an assessment of Nagel's view, see chapter 7.

'understanding something to be a complete thing' (which concerns solely the nature of my *understanding*) and 'a complete understanding of something' (which seems to concern also the nature of the *thing* conceived) *to have the same meaning* (CSM 2, Rep 4: 156). However, even if my conception of *y* appears to be complete (given the usual provisos concerning clearness and distinctness), for it to *be* complete it must be such that the attribution of *F* to *y* (or the denial of *Fy*) is compatible with the essence of *y*. This much seems to be clear from Descartes' claim that

> the idea of a substance with extension and shape is a complete idea, because I can conceive it entirely on its own, and [positively] deny of it everything else of which I have an idea. (CSMK: 202)

With some elaboration, it seems as if the following holds: a truly adequate conception of *y* is such that all essential properties of *y* are (in some sense) *explicit* to the conceiver. By contrast, a truly complete conception of *y* is such that all essential properties of *y* are *implicit* in the conception (if not, how could one know that the attribution of a certain property *F* to *y* is compatible with the essence of *y*?). Now, I *understand my mind to be a complete thing*. But for *myself to have a complete conception of my mind*, my conception must be such that for any property I find myself able to attribute to my mind given this conception, that property must be compatible with the essence of my mind. It is easy to see how the Arnauldian skeptic could draw a distinction between apparently and genuinely complete conceptions given these interpretations of the notions involved. We could say that a conception of *y* is *apparently* complete if *y* is understood to be a complete thing, that is, if *y* is understood as a thing that is capable of independent existence. By contrast, we could say that a conception of *y* is *genuinely* complete if it *is* complete, that is, if *y* is, as conceived, a thing that *is* capable of independent existence.

In summary, the merit of the amended separability argument seems to be that Descartes replaces Arnauld's demand for adequate conceptions with the requirement for complete conceptions. Admittedly, the requirement for complete conceptions is—at least according to our interpretation—a weaker requirement. In this sense, the separability argument has been improved in so far as it puts less requirements on the conceiver. However, it seems as if the problems with the argument remain the same, and that Arnauld's initial criticism still applies when spelled out in different terms.

When it comes to Yablo's attempt to shift the burden of proof back to those who are skeptical about the value of conceivability intuitions, we shall have to return to it below (see chapter 7). Yablo's argument should be assessed in a

more general discussion of the evidential value of conceivability intuitions, since his ideas concerning the "burden of proof" do not, as such, pertain exclusively to Descartes' argument, but rather to any argument proceeding from conceivability claims.

4.3 Is there a limit to Arnauldian skepticism?

After having reviewed the exchange between Descartes and Arnauld, one might wonder whether there is any limit to Arnauldian skepticism. Are there any modal claims that are safe from doubt? There are two ways in which Arnauldian doubt is restricted. First, let us extrapolate from what Arnauld grants Descartes in his objection, and take the following general thesis to be true:

(12) If *x* believes or (honestly) claims that it appears to *x* that S, then it appears to *x* that S.

Tye (1995) argues that it is absurd to doubt (12). Could it be, he asks, "that I merely believe that I am seeming [or that I *merely appear*] to see chairs, desks, trees, people, and so on, when in reality *I am not even seeming to see* these things?" (1995: 192, my emphasis). He answers: "It seems to me that we have a strong, pretheoretical conviction that error of this sort is absolutely impossible. It is simply absurd to suppose that I, or we, could be wrong in these ways" (1995: 192). I agree, and there does not seem to be any reason why we cannot extend this principle to cover appearances of conceivability or possibility. *Mutatis mutandis*:

(13) If *x* believes or (honestly) claims that it appears to *x* that *x* can conceive that S, then it appears to *x* that *x* can conceive that S, that is, *x* has *an apparent conception* that S.

Secondly, the basis for a more interesting limit to Arnauldian skepticism can be found in Arnauld's correspondence with Leibniz. Leibniz held that the essence of a thing includes *all* its properties (Yablo 1990: 161), as is reasonably evident from his claim that

> if in the life of some person and even in this entire universe something were to proceed in a different way from what it does, nothing would prevent us saying that it would be another person or another possible universe that God would have chosen. It would thus truly be another individual; there must also be an *a priori* reason (independent of any experience) which makes one say truly that it is I who was in Paris and that it is still I, and not another, who am now in Germany, and

> consequently *the concept of myself must link or include the different states*. (LAC: 59–60)[11]

As Yablo (1990: 161–2) points out, Arnauld objects that he could decide not to make a certain journey, and this would not prevent him from being himself. (On Leibniz' view, a person who actually went from Paris to Germany on a certain occasion would not be the "same" person if she, contrary to fact, decided to stay in Paris.)

Arnauld's objection to Leibniz reveals a second limit to Arnauldian skepticism. Yablo argues that we should distinguish between the *Arnauldian skeptic*, the philosopher who doubts that Descartes is entitled to conclude that his mind is distinct from his body, and the *Arnauldian believer*, the philosopher who rejects the claim that in the scenario where I decide (counterfactually) not to make a certain journey, I am not the same person as I actually am (cf. Yablo 1990: 161–2). In what follows, we shall encounter philosophers that seem to be *both* Arnauldian skeptics and Arnauldian believers. In particular, I take van Inwagen (1998) to promote precisely such a view. Yablo argues that you could press the Arnauldian believer at the point where he implicitly claims to know that *making the Paris-Germany trip* is not part of his essence. It is hard to see, Yablo argues, what it is that entitles the Arnauldian believer to *this* conclusion, when he, for example, rejects the separability argument (cf. Yablo 1990: 162). "To be consistent," Yablo has repeatedly pointed out, "Arnauld should hold that *all de re* conceivability intuitions are suspect, unless the ideas employed are certifiable in advance as adequate" (1990: 159, also with minor reformulations in 1993: 16). However, Yablo argues that there must be some way of escaping this radical form of modal skepticism.

---

[11] The beginning of this passage is also cited in Plantinga 1979 and Yablo 1990. See further Leibniz' claim that "each individual substance is an expression of the entire universe […] [and] in its concept all events that occur in it are included with all their circumstances and the whole succession of external things" (LAC: 5), and the (equivalent?) claim that "the individual concept of each person contains once and for all everything that will ever happen to him" (LAC: 53). From these claims, it may be taken to follow that Leibniz view is that each object only exists in one possible world (Plantinga 1979: 147–8). On Leibniz' view, the assumptions (i) *a* exists in more than one possible world, say in *w* and *w**, and (ii) *w* and *w** are *distinct* possible worlds, are inconsistent. For two possible worlds to be distinct, there must be some fact that obtains in one of them but not in the other. However, if the worlds are distinct in this sense, regarding *any* fact, then they will be different with respect to facts about *a*, since on Leibniz' view *all* facts that obtain in the universe of *a* are also facts about *a*. But then *a* cannot exist in more than one possible world since it cannot be different in any respect in so far as it would then be *another object*.

> If the [Arnauldian] skeptic's doubts are allowed to stand, then it is not obvious how the [Arnauldian] believer can hope to refute Leibniz' suggestion that my essence takes in all my properties whatsoever! Yet surely we side here with the believer. Even without an answer to the skeptic, I think we feel that he *must* be wrong. Somehow or other, I *must* be in the position to refute the suggestion that I am essentially born on the day of my actual birth, or, even more unbelievably, essentially surrounded by the entire course of actual history. (1990: 162)

Evidently, then, some philosophers (including, apparently, Arnauld himself) have seen a principled difference between modal statements such as "I could have decided not to make the trip from Paris to Germany" or Yablo's "I could have been born on another day," or van Inwagen's "The table at which I sit could have been two feet to the left" (cf. 1998: 75–6), on the one hand, and modal statements such as the conclusion of Descartes' mind-body argument, on the other. The former seem to be immune to Arnauldian doubt, whereas the latter are clearly subject to such doubt. Wherein lies the difference? One possible way to answer this question has been suggested by van Inwagen (see sections 2.5–2.6 and 3.3 above): if we restrict modal judgments to states of affairs which are, in some sense, basic and familiar, these judgments are exempted from Arnauldian doubt. In chapter 7, we shall return to these questions.

4.4 Contemporary mind-body arguments

In this section I shall present two contemporary mind-body arguments, the first of which is due to Kripke (1980) and the second to Chalmers (1996). Both of these arguments are similar to Descartes' separability argument in certain respects. However, neither Kripke nor Chalmers attempts to establish the same conclusion about the relation between mind and body as Descartes. Descartes attempts to show that mind and body are different substances (or that the mind is distinct from the body). Kripke and Chalmers, on the other hand, do not attempt to argue for the distinctness of the mind *as a whole* from the body *as a whole*. Instead, they focus on a class of experiences that are associated with a certain *aspect* of the mind. At least Chalmers makes a sharp distinction between the *psychological* and the *phenomenal* aspect of mind.[12] In general, Chalmers says, if one describes mental states in terms of what they *do*, one is concerned with

[12] Chalmers notes that Descartes was prone to conflate the psychological aspect of the mind with its phenomenal aspect (Chalmers 1996: 12).

the psychological aspect of mind.[13] On the other hand, if one describes mental states in terms of *how it feels* (or *what it is like*) to be in them or have them, then one is concerned with the *phenomenal* aspect of mind (Chalmers 1996: 11). Kripke's and Chalmers' arguments concern the relation between (manifestations of) the latter aspect of mind and the physical world.[14] Moreover, both Kripke and Chalmers try to anticipate the criticism that may be made by an Arnauldian skeptic. Perhaps this is the foremost virtue of their arguments.

4.4.1 Kripke

Kripke's *Naming and Necessity* (1980) is perhaps the most influential contemporary attempt to bring Cartesian and Arnauldian ideas about conceivability and possibility together. Kripke provides a detailed analysis of the ways in which conceivability intuitions can be confused and misleading, and his considerations are similar to Arnauld's objections to Descartes. However, it is clear that Kripke is not skeptical towards *all* modal judgments. Even if the Arnauldian skeptic is correct in his claim that only conceivability based on adequate conceptions implies possibility, says Kripke, adequate conceptions are not *necessary* in order to establish that something is *impossible* or *necessary* (cf. 1980: 46). In order for us to know that it is *possible* for, say, Nixon to be an inanimate object, we must, according to Arnauld, know that the property of being inanimate is compatible with all of Nixon's essential properties taken together. This, in turn, requires that we know all of Nixon's essential properties, and that we know that he has no essential properties apart from the ones we know of. However, Kripke argues, in order to know whether the possession of a certain property *F* is a *necessary* requirement for "Nixonhood" (i.e., for being *Nixon*), we do not need to (be able to) spell out purely qualitative *sufficient* conditions for Nixonhood (cf. 1980: 46) (That is, we do not need to know *all* of Nixon's essential properties.) In order to know whether the possession of a certain property *F* is a *necessary* requirement for Nixonhood, we should instead attempt to imagine a possible world in which Nixon does not have the property *F*. If we cannot imagine such a world, then *F* is a necessary property of Nixon, or, equivalently, a necessary requirement on Nixonhood (cf. 1980: 46). Thus, he

---

[13] Chalmers 1996: 11. The psychological aspect of mind involves, according to Chalmers, *awakeness*, *introspection*, *reportability*, *self-consciousness*, *knowledge*, and so forth.

[14] There are no reasons to suspect that Kripke would not approve of Chalmers' attempt to distinguish between the psychological and the phenomenal aspect of mind.

thinks that we can obtain knowledge of what is *impossible* and *necessary* for Nixon without knowing all of Nixon's essential properties.[15]

However, when it comes to conceivability (or imaginability) as a guide to what is *possible* for a certain object, Kripke's ideas include a certain amount of Arnauldian skepticism. Kripke's general contention is that conceivability does imply possibility.[16] However, he argues, we sometimes *misdescribe* the scenario that we actually conceive of (the scenario $\mathcal{S}$*) as being another (qualitatively indistinguishable) scenario $\mathcal{S}$, to the effect that we infer the possibility of $\mathcal{S}$ (which we sought to conceive of) rather than $\mathcal{S}$* (which we actually conceived of).[17] To understand what this means, we begin by considering what it means for two conceivers to be in what Kripke calls *qualitatively indistinguishable epistemic situations* (1980: 103–4).[18] Consider a qualitative duplicate of our world, Twin World. In Twin World, some empirical facts differ from the facts about our world, but they differ in ways that do not affect the phenomenal nature of things. With respect to the phenomenal nature of things, our world and Twin World are indistinguishable, at least to the naked eye. Now, one of the facts about Twin World that differs from the facts about our world is that in Twin World the stuff that fills the lakes is not H_2O but XYZ—a substance which is qualitatively indistinguishable from H_2O. Now consider my twin in Twin World. When he goes to the tap for a glass of what he calls "water," he is in a situation that is *qualitatively indistinguishable* from my situation when I go to the tap for a glass of water. Nevertheless, what comes out of his tap (XYZ) is something other than what comes out of mine (H_2O). Now, suppose further that the chemical composition of the stuff that comes out of my twin's tap is unknown to him, and that the chemical composition of the stuff that comes out of my tap is unknown to me. If so, my twin and I are in two qualitatively indistinguishable epistemic situations.

Now, Kripke argues, it seems conceivable that water—the stuff that comes out of my tap—could have been something other than H_2O, because there is a qualitatively indistinguishable epistemic situation in which the stuff that comes out of the tap is something else than H_2O, namely the situation of my twin. Thus, I can simply imagine myself being identical to my twin, to the effect that *I*

---

[15] See section 5.4.1 and chapter 6 below for further discussion regarding these central claims in Kripke 1980: 46–7.

[16] Cf. Gendler and Hawthorne 2002: 38; Yablo 1999: 45–56.

[17] Cf. Yablo 1999; Gendler and Hawthorne 2002: 34

[18] Kripke's explanation is reminiscent of Putnam's Twin Earth thought-experiment. See Putnam 1975.

am in a situation which is qualitatively exactly the same as my actual situation—in which the stuff that comes out of the tap and fills the lakes is not H_2O. However, Kripke argues, this is *not* to imagine a scenario in which *water* (again, the stuff that *actually* comes out of my tap) is not H_2O. The imagined scenario is simply a scenario where something—"the stuff that comes out of the tap"—that has all the appearances of water, is not H_2O. In short, when we find it conceivable that water could have been something else than H_2O, we *misdescribe* the scenario that we have *actually conceived* of (the scenario $\mathcal{S}$*, where something with all the appearances of water is not H_2O) as a scenario where *water* is not H_2O (that is, the scenario that we *sought* to conceive of, $\mathcal{S}$). As Kripke remarks:

> The loose and inaccurate statement that gold might have turned out to be a compound should be replaced (roughly) by the statement that it is logically possible that there should have been a compound with all the properties originally known to hold of gold. The inaccurate statement that Hesperus might have turned out not to be Phosphorus should be replaced by the true contingency […]: two distinct bodies might have occupied, in the morning and the evening, respectively, the very positions occupied by Hesperus-Phosphorus-Venus. (1980: 142–3)

The question now is as follows. How can we know, concerning a particular inference from conceivability to possibility, that the scenario we conclude is possible is the same scenario as the scenario we actually conceived of? Maybe we are subject to the same confusion as in the water/H_2O case, where we mistook the conceivability of $\mathcal{S}$* for the conceivability of $\mathcal{S}$? Above, we reviewed Arnauld's attempt to avoid such mistakes according to which conceptions should fulfill certain criteria in order to be admissible. Kripke, on the contrary, never considers imposing requirements on *conceptions*. Instead, he identifies a class of phenomena that seem to be immune to the type of confusion in the water/H_2O case. This class includes (at least) our phenomenal experiences. For example, when I imagine a scenario in which I am in pain, I cannot be subject to the same confusion as in the water/H_2O case, because I cannot be mistaken about what it is that I imagine. It cannot be that I imagine a scenario in which I am beset with something with all the appearances of pain but in which I am nevertheless not *in pain*. Any scenario which is qualitatively indistinguishable from a scenario in which I am in pain must itself be a scenario in which I am in pain (cf. McGinn 1976: 79).

Given these ideas, the path to the mind-body argument is straightforward. Let P be a specific pain experience and let B range over physical (brain) states. Now, Kripke argues that for each particular B, it is conceivable that P should exist without B. (That is, for each B there is a conceivable scenario $\mathcal{S}$ in which

P exists without *B*.) But then it follows from the conceivability thesis and the logic of identity that *P* is distinct from *B* for each particular *B*. In other words, there are facts about phenomenal mental states that cannot be reduced to (or identified with) facts about physical states (of the brain). If this is true, materialism (the doctrine that all entities are composed of, or reducible to, matter, material forces or physical processes or states[19]) must be false.

The crucial step in the argument is, again, the inference from conceivability to possibility. However, as we have seen, Kripke thinks that in this particular case the argument is immune to Arnauldian criticism. In this particular case, apparent and genuine conceivability coincide (cf. Kripke 1980: 152–3).

4.4.2 Chalmers

An argument that is similar to both Descartes' and Kripke's arguments is Chalmers' *zombie argument*, which we briefly reviewed in section 2.7 above. The zombie argument is directed against the materialist claim that any possible world that is a physical duplicate of the actual world is a duplicate of the actual world *simpliciter* (cf. Chalmers 1996: 38; 41–2). Using Kripke's account, this thesis can be understood as equivalent to the claim that once God had determined all actual physical facts, he did not, according to the materialist, have to *do any additional work* in order to make the actual facts about consciousness hold (cf. Chalmers 1996: 38; Kripke 1980: 153–4).

Chalmers' argument starts from the claim that a *zombie world* is ideally conceivable and therefore logically possible (Chalmers 1996: 96–9; 1999a: 477n; 1999b: 436). In order to understand to what this claim amounts, some concepts must be clarified (these concepts will be subjected to further study in chapter 5): First, a *zombie world* is, by definition, a world that is an exact physical duplicate of our world, but where there are no conscious phenomenal experiences. The inhabitants of the zombie world—the zombies—are thus physical duplicates of us that lack phenomenal experiences.[20] Secondly, a statement S is ideally conceivable iff S is conceivable on *ideal rational reflection* (2002a: 147).[21] Chalmers goes on to define "ideal rational reflection" in terms of *undefeatable reasoning*. Given the notion of undefeatable reasoning, Chalmers says, "we can say that S is ideally conceivable when there is a

---

[19] Cf. "Materialism," in Craig, E. (ed.), 1998. *Routledge Encyclopedia of Philosophy*, volume 6. London: Routledge.

[20] Chalmers further defines zombies as *functional* and *psychological* duplicates of us (1996: 95).

[21] Chalmers takes conceivability to be a property of statements. See Chalmers 2002a: 147.

possible conceiver for whom S is *prima facie* conceivable [that is, conceivable after some consideration], with justification that is undefeatable by better reasoning" (2002a: 148). Finally, on Chalmers' usage, a statement is "logically possible" iff S is broadly logically (or *conceptually*) possible in the sense outlined in section 2.3 above (1996: 131; 1999a: 477). In other words, Chalmers' claim is that a zombie world is conceivable on undefeatable rational reflection over the concepts involved in the description of such a world, and it is therefore broadly logically, or conceptually, possible.

Given Chalmers' view, it is prima facie reasonable to take ideal conceivability to imply logical possibility, as Chalmers does. Suppose that a conceiver whose reasoning is undefeatable by better reasoning is unable to detect any conceptual contradiction in a statement S—or even better, that this conceiver can clearly "see" that S involves no such contradiction—upon examination of the concepts involved in S (this is the interpretation of what it means that something is conceivable that will be relevant below). If this is the case, Chalmers argues, then this provides a very strong reason to think that S is logically possible (cf. Chalmers 1996: 99). Hence, it also seems that *given* that a zombie world is conceivable (as described), then such a world is also logically possible (as described). But if a zombie world is logically possible, Chalmers argues, the materialist claim that any physical duplicate of the actual world is a duplicate of the actual world *simpliciter* must be wrong. In the actual world, certain facts about phenomenal consciousness obtain, but these facts *do not* obtain in the zombie world. For example, there is something it is like for me to sense the smell of coffee, but there is nothing it is like for my zombie counterpart to sense the smell of coffee. In other words, if we again use Kripke's account, God *had* to do some additional work in order to make the physical world *our* world instead of the zombie world: he had to connect the physical stimulation caused by the coffee with the relevant characteristic phenomenal experience (cf. Chalmers 1996: 124). (In order to create the physical world, it is thought, God only had to create the *zombie world*, and in that world there are no phenomenal experiences. In particular, there are no phenomenal experiences associated with the physical stimulation caused by coffee.)

There are a number of ways in which one may object to this argument. Some authors have questioned the conceivability thesis, which supports the inference from conceivability (in the sense described) to possibility (in the sense described). (Such general skepticism towards conceivability intuitions has already been addressed in a more general setting, and will be addressed again, below.) For example, as we saw above, the central idea of Churchland (1998)

and Nagel (1998) is that our apparent ability to conceive of zombies is due to our psychological limitations, or to our limited understanding of the processes involved. Such limitations in our current conceptual scheme will make inferences from conceivability to possibility susceptible to doubt. Other authors have not explicitly questioned the conceivability thesis, but they deny that a zombie world is coherently conceivable (Shoemaker 1999: 440, 442), and that, in effect, what has actually been conceived is *misdescribed* as a zombie world (cf. Chalmers 1996: 99; see also section 4.4.1 above). Another view admits that the argument may suffice to establish that a zombie world is *logically* possible. However, it is claimed, the argument does not establish that a zombie world is *metaphysically* possible, which is what is needed in order to refute the materialist thesis; after all, it is argued, it is "logically possible" that water could have been something other than H_2O, but, nevertheless, the world could not have been that way (cf. Chalmers 1996: 131).

In response to the first objection, Chalmers points to the distinction between *prima facie* and *ideal* conceivability, in terms of which his argument is formulated. This distinction will be dealt with more thoroughly later on, but I shall indicate the main line in Chalmers' answer. As we have seen, Chalmers' argument involves the claim that *ideal* conceivability implies possibility. Chalmers proposes that we reinterpret Churchland's and Nagel's criticism as follows: a zombie world is *prima facie* conceivable, but it is not *ideally* conceivable (cf. Chalmers 2002b). Chalmers' proposal amounts to the following: Churchland and Nagel argue that a zombie world is conceivable given our present state of knowledge and our present conceptual scheme, but a zombie world would not be conceivable on ideal grounds, that is, on ideal rational reflection (which may also involve a sufficient understanding of the relevant physical facts and conceptual innovations). Chalmers, on the other hand, argues that there is no hidden inconsistency involved in the description of a zombie world, and therefore maintains that a zombie world *is* ideally conceivable. To the outside observer, it seems as if a stalemate has occurred between Chalmers and his critics at this point.

In response to the second objection, Chalmers challenges the objector to point to the supposed inconsistency in the description of the zombie world. (Compare Yablo's attempt to shift the burden of proof to the Arnauldian skeptic in section 4.2.2 above.) Chalmers says:

> Perhaps if we thought about it [that is, a zombie world] clearly enough we would realize that by imagining a physically identical world we are thereby *automatically* imagining a world in which there is conscious experience. But then the burden is on the opponent to give us some idea of where the contradiction

> might lie in the apparently quite coherent description. If no internal incoherence can be revealed, then there is a very strong case that the zombie world is logically possible. (Chalmers 1996: 99)

Central to Chalmers' response to the third objection (which draws on the distinction between logical and metaphysical possibility) is his claim that the distinction between logical possibility and metaphysical possibility only applies to *statements*, not to *worlds*. According to Chalmers, there is only one space of possible worlds: the space of logically possible worlds (1996: 134 and *passim*). Some of the logically possible worlds may be *misdescribed* (in the sense outlined in the previous section) by confusing what Chalmers calls *the primary intension* of a certain term (which roughly corresponds to a Kripkean *reference-fixing description*: "the watery stuff") with the *secondary intension* of the same term (which is basically the reference of the term in our world: "H_2O").[22] For example, a logically possible world in which *the watery stuff* is XYZ may be misdescribed as a world in which *water* is XYZ.

The distinction between primary and secondary intension on the statement level can be explicated as follows. Consider the following statement:

(14) Water $\neq H_2O$.

What is stated in (14) in our world is *false* since 'water' and 'H_2O' refer to the same substance, namely, the watery stuff that comes out of the tap and fills the lakes. By contrast, (14) expresses a *truth* when uttered in Twin World, where the stuff in the lakes is also called 'water', but where that stuff is XYZ. In accordance with my exposition in section 1.3 above and these distinctions, I shall call *what is said* or *expressed* by a certain statement S when uttered in a world *w* the *proposition* expressed by S in *w*. I further define a *propositional function* $[\![S]\!]$ associated with a statement S to be a function which, for each possible world *w*, yields the proposition that S would have expressed had *w* been the actual world (or, the proposition that S would have expressed if uttered in *w*). Thus, $[\![S]\!]_@$ (where we take '@' to be a name of the actual world) is the proposition actually expressed by S. In general, $[\![S]\!]_w$ is the proposition that S would express had S been uttered in *w*. For example, $[\![\text{Water} \neq H_2O]\!]_@$ is the proposition that $H_2O \neq H_2O$, and $[\![\text{Water} \neq H_2O]\!]_{\text{Twin World}}$ is the proposition that $XYZ \neq H_2O$.

Now, we say that a statement S is *possible with respect to its primary intension*, or *primarily possible*, if there is at least one logically possible world *w* such that the proposition $[\![S]\!]_w$ is true in *w*. Now, is (14) primarily possible? The

---

[22] Cf. Chalmers 1996: 132.

answer is yes, since there are many worlds *w* such that $[\![\text{Water} \neq H_2O]\!]_w$ is true in *w*. One of them is Twin World.

Let us now fix the world *w* of the propositional function as *our world*, @, where the watery stuff is H_2O. We say that a statement S is *possible with respect to its secondary intension*, or *secondarily possible*, if there is at least one logically possible world *w* such that the proposition $[\![\text{Water} \neq H_2O]\!]_@$ is true in *w*. For example, consider (14) again. The proposition $[\![\text{Water} \neq H_2O]\!]_@$ is a *necessarily false* proposition, that is, a proposition that is false in all logically possible worlds, namely that $H_2O \neq H_2O$. Thus, there is no logically possible world in which the proposition $[\![\text{Water} \neq H_2O]\!]_@$ is true, and (14) is thus not secondarily possible.

It is important to note that the above distinctions are made relative to one space of possible worlds—the logically possible worlds—and that the distinction between primary and secondary possibility strictly pertains to how a certain statement S is evaluated. Given this, Chalmers' reply to the third objection can be spelled out as follows. The third objection says that many statements are logically but not metaphysically possible, for example, "Water ≠ H_2O," and "Hesperus ≠ Phosphorus." By implication or analogy, might it not be that zombies are logically possible but metaphysically impossible? As we have seen, according to Chalmers', there is no room for a distinction between logical possibility and metaphysical possibility on the level of worlds—there is just one set of worlds, the logically possible ones. According to him, the attribution of *logical* possibility to a statement S comes down to the truth of $[\![S]\!]_w$ in *w*, for some logically possible world *w*. Correspondingly, the *metaphysical* possibility of a statement S comes down to the truth of $[\![S]\!]_@$ in *w*, for some logically possible world *w* (cf. Chalmers 1996: 132).

Now, let Z be the description of a zombie world. Chalmers concedes that for some logically possible world *w*, $[\![Z]\!]_w$ *may* be a proposition that *differs* from the proposition $[\![Z]\!]_@$, in the same way as the proposition $[\![\text{Water} \neq H_2O]\!]_{\text{Twin World}}$ differs from the proposition $[\![\text{Water} \neq H_2O]\!]_@$ (cf. 1996: 132). In other words, he concedes that the primary intensions, or "reference-fixing descriptions," of the terms involved in the description of the zombie world may differ from the secondary intensions of the same terms, in the same way as the primary intension of 'water' ("the watery stuff") differs from the secondary intension of the same term ("H_2O").[23]

---

[23] Cf. Chalmers 1996: 132. Kripke holds, as we saw above, that such a difference is not possible when it comes to the *phenomenal concepts* employed in the description of the

Chalmers' reply to his objectors is that even if there is such a difference this is irrelevant to the zombie argument. In his presentation of the argument, Chalmers focuses on the *what it is like* component of conscious experiences, and argues that a world in which all physical facts are the same as in our world, but in which there are no phenomenal experiences (or "what it is like"-experiences) is conceivable and therefore logically possible. The different states of the *phenomenal* mind are identified with, *and by*, "what it is like"-descriptions. Even if there is something to the relevant (total) states of mind that is experientially inaccessible to us (something which is picked out by the *secondary* intensions of the terms involved in the description of a zombie world), the logical possibility of a world which is physically like ours but devoid of "what it is like"-experiences suffices for the argument to be valid, Chalmers argues:

> For note that whether or not the primary and secondary intensions coincide [with respect to a statement S], the primary intension determines a perfectly good property of objects in possible worlds. The property of being watery stuff is a perfectly reasonable property, even though it is not the same as the property of being H_2O. If we can show that there are possible worlds that are physically identical to ours but in which the property introduced by the primary intension is lacking, then dualism will follow.
>
> This is just what has been done with consciousness. We have seen that there are worlds physically just like ours that lack consciousness, according to the primary intension thereof. This difference in worlds is sufficient to show that there are properties of our world over and above the physical properties. By analogy, if we could show that there were worlds physically identical to ours in which there was no watery stuff, we would have established dualism about water just as well as if we had established that there were worlds physically identical to ours in which there was no H_2O. (Chalmers 1996: 132–3)

I believe that there is a sense in which Chalmers' reply to the third objection is similar to Descartes' reply to Arnauld. At least, Descartes' claim that although he may have many properties of which he is not yet aware, he is certain that he could have been created by God without having the properties of which he is *unaware*, seems to be analogous to Chalmers' claim that the logical possibility of a world wherein the (directly perceptible) "what it is like" properties are lacking—disregarding the other properties—suffices to support dualism. We shall return to this idea in section 7.4 below.

---

zombie world. The "reference-fixing description" used to determine the reference of 'pain', for example, *necessarily* picks out pain.

4.5 Summary

The purpose of this chapter has been twofold. First, we have attempted to problematize the relation between conceivability and possibility by looking at how the conceivability thesis is employed in philosophical arguments, and how these arguments have been criticized. The second purpose of this chapter has been to suggest that the strong interpretation of what it means that something is possible that we introduced in chapter 3 is employed in each of these arguments.

First, it is clear that Descartes' mind-body argument, and the exchange between Descartes and Arnauld, concerns how the world could have been, and not what is possible in some weaker sense. Although I take this much to be evident, I shall make two further points that support this conclusion. First, Descartes took conceptual possibility to coincide with strict possibility. Because God can make real everything that is conceivable, there is, in Descartes, no distinction between conceptual and strict possibility. It is true that Descartes acknowledges that mistaken modal judgments may occur. However, such mistakes occur when we mistakenly take some statement to be a conceptual (and therefore strict) possibility because our conception is not clear and distinct. Therefore, the fact that Descartes admits that modal mistakes occur does not involve any distinction between conceptual and strict possibility; rather, it pertains to our abilities to recognize conceptual possibilities and necessities. Secondly, from the fact that

(1) What I can conceive clearly and distinctly is possible,

Descartes infers (by means of the thesis that all *possibly* distinct things are *actually* distinct) what is *actually* the case (in the world):

(2) If I can conceive clearly and distinctly of x as being distinct from y, then x is distinct from y.

If Descartes had employed a concept of possibility that was weaker than that of strict possibility, the inference from what could have been the case to what is actually the case would have been invalid.

With respect to Kripke's argument, I also take it to have been shown that the notion of possibility involved in that argument is that of strict possibility.[24] Here

[24] As we introduced the notion of strict possibility in chapter 3 above, we noted that the concept of strict possibility is similar to Kripke's concept of *metaphysical* possibility, in that both are defined in terms of how the world could have been. Our reasons for not adopting the Kripkean notion of metaphysical possibility is that we wanted our notion to be, at least initially, independent of any theories about *what type of facts* determines how the world could

we should emphasize that one of the objectives of *Naming and Necessity* is to introduce a notion of possibility in terms of how the world could have been, and to investigate various philosophical problems (concerning the reference of singular terms, mind-body identity, and so forth) in relation to that notion. Kripke is the modern champion of the distinction between apparent and genuine conceivability, and he takes genuine conceivability to be determined by strict possibility, in the sense that nothing is genuinely conceivable unless it is possible.[25] Many arguments in *Naming and Necessity* attempt to demonstrate how the apparent conceivability of some scenario $\mathcal{S}$ is actually a conception of a qualitatively similar scenario $\mathcal{S}$*, where the conceivability of $\mathcal{S}$* is mistaken for the conceivability of $\mathcal{S}$. According to Kripke, it is true that if $\mathcal{S}$ is conceivable, then $\mathcal{S}$ is possible, but the fact that $\mathcal{S}$ is *apparently* conceivable—that is, that $\mathcal{S}$* is genuinely conceivable—does not tell us anything about the possibility of $\mathcal{S}$. However, Kripke holds that apparent and genuine conceivability coincide with respect to our subjective experiences. With respect to these, there is no danger of confusing the actual object of conception with some other object. Thus, as regards our phenomenal experiences, we can safely infer strict possibility from conceivability.

Finally, with respect to Chalmers' argument, it suffices to point out that the main point of this argument seems to be that phenomenal properties, that is, the "what it is like" components in our experiences, taken by themselves, are perfectly good properties of things in other possible worlds. According to Chalmers, the distinction between logical and metaphysical possibility is a distinction between the *primary* intension of some statement being true in some world and the *secondary* intension of some statement being true in some world. However, both notions are concerned with *truth in some possible world*, to wit, with strict possibility, or with ways in which the world might have been.

---

have been. (It may be that the notion of metaphysical possibility is associated with such theories.)

[25] As Almog (2002a: 43) puts it: "In Kripke, as in Arnauld, the epistemological is reined by the metaphysical."

CHAPTER 5

INTERPRETATIONS OF CONCEIVABILITY

5.1 Introduction

In chapter 3, I presented three main questions regarding the conceivability thesis. First, what shall we *take it* to mean that something is *conceivable*? Secondly, how shall we interpret 'implies' in "conceivability implies possibility"? Finally, what kind of possibility shall we take the conceivability thesis to attribute to conceivable things?

In chapter 3, I also presented tentative answers to the two latter questions. First, I considered whether a strong interpretation of 'implies' is viable, according to which conceivability is sufficient evidence for possibility. Secondly, I argued for an objective notion of *strict possibility*, which I hold to be the notion of possibility employed in certain central philosophical arguments (which I discussed in chapter 4). On this notion of possibility, "S is possible" (for any statement S) should be interpreted as follows: S is possible if the world could have been such that S was true, or, if there is a *possible world* such that S would have been true had that world been actual. This analysis in some sense commits us to the existence of possible worlds. However, at the present stage of analysis, our account is intended to be compatible with any theory about the nature of possible worlds.

In the present chapter, I shall focus on the question of what it means—or what we shall take it to mean—that something is conceivable. I shall distinguish between various *types* of distinctions one can make concerning the notion of conceivability, and I shall consider various proposals concerning how to interpret 'conceivable' relative to these distinctions. The present chapter does not attempt to provide a complete list of all the definitions of conceivability that have actually been presented (although my exposition, I believe, borders on being complete in this respect). Nor do I attempt to categorize all the definitions of conceivability under consideration with respect to the distinctions made. The goal is rather to provide a comprehensive overview of the alternative options one is faced with when defining conceivability. I shall then pursue the options that seem most viable. To this end, I shall close the present chapter by

presenting a number of desiderata that I argue that any definition of conceivability should fulfill (given the goals pursued in the present thesis), and I take a stand on central issues regarding the question of which desiderata should be given prescedence.

Prima facie, one can make at least four different types of distinctions with respect to the notion of conceivability. First, there is the distinction between views affirming *absolute* and *relative* conceivability. On views of the former kind conceivability is a *property* of entities such as statements, propositions, states of affairs, and so on. By contrast, on views of the latter kind conceivability is a *relation* that may or may not obtain between entities of the aforementioned kinds and a conceiver. According to one notion of conceivability, a statement S is conceivable for a conceiver *a* iff S is epistemically possible for *a*, in the sense of being compatible with everything *a* knows. This interpretation is set out *relative* to a conceiver, and is therefore an example of a *relative* notion of conceivability.

Secondly, there are distinctions that pertain to *the manner in which something is conceived*. Obviously, these distinctions are only relevant given a *relative* interpretation of conceivability. The literature suggests that one can distinguish (at least) between *positive* and *negative*, *primary* and *secondary*, and *ideal* and *prima facie*, modes of conception (Chalmers 2002a). Consider, for example, the distinction between *positive* and *negative* conceivability.[1] As Chalmers notes, positive interpretations of conceivability typically require that a conceiver *succeeds* in performing some mental act, such as forming a mental image of a certain state of affairs, describing a scenario, etc. A *negative* interpretation, on the contrary, typically requires the *absence* of something. For example, a negative interpretation of what it means that something is conceivable may be that S is conceivable for a conceiver *a* iff *a* cannot intuitively see that S is impossible. It is clear that the latter account involves a requirement that something should *not* obtain (namely, *a*'s ability to "see" that S is impossible).

Thirdly, one can distinguish between different *substantial* notions of conceivability. By a "substantial" notion of conceivability, I mean the component of any complete notion of conceivability that explains in cognitive terms what must be the case with respect to a conceiver in order for him or her to conceive of something. For example, Yablo (1993) defines conceivability as follows: a statement S is conceivable for a conceiver iff the conceiver can imagine a world *w* such that the conceiver takes S to be true in *w*. Yablo's definition of

[1] Variants of this distinction can be found in Van Cleve 1983, Yablo 1993, and Chalmers 2002a. See below.

conceivability thus incorporates an account of what is required in terms of the mental state of a conceiver in order for that conceiver to be conceiving of (the truth of) a statement S. First, the conceiver must imagine something—a possible world—and secondly, the conceiver must entertain a certain attitude to the content or object of his or her imagination—he or she must believe the imagined world to be a world in which S is true.

Fourth and finally, there is the distinction between *apparent* and *genuine* conceivability, which we have discussed extensively in previous chapters. It is debatable whether this distinction only pertains to the concept of conceivability itself. Rather, as it is often used, the distinction concerns the relation between conceivability and possibility. Accounts of the difference between apparent and genuine conceivability normally distinguish one sense of "S is conceivable," according to which conceivability does not imply possibility, from another sense according to which conceivability does imply possibility. (See, for example, the discussion of Arnauld's ideas in section 4.2.1 above.) An account of the distinction can make use of two different interpretations of conceivability. Alternatively, such an account may discern different modes of conception, given the same substantial notion of conceivability. As an example of the latter, suppose that our notion of conceivability is that a statement S is conceivable for a conceiver just in case the conceiver cannot detect any (conceptual) contradiction in S. We may then stipulate that S is *genuinely* conceivable if an ideal conceiver, aided by unlimited powers of reason, cannot detect any contradiction in S (this notion is akin to the notion of *ideal primary conceivability* in Chalmers 2002a). By contrast, we can say that S is *apparently* conceivable iff a normal, limited conceiver (a conceiver that possesses the same powers of reason as an ordinary human being) cannot detect any contradiction in S on first appearances (this notion roughly corresponds to the notion of *prima facie conceivability* in Chalmers 2002a). Suppose now that our concept of possibility is (a variant of) *conceptual possibility*, according to which S is possible iff S does not, in fact, involve any logical or conceptual contradiction. Then it is reasonable to suppose that *genuine* conceivability (thus outlined) implies possibility, but that *apparent* conceivability (thus outlined) does not.

The main question when it comes to the distinction between apparent and genuine conceivability is whether such a distinction can be allowed. In chapter 4, I have outlined the distinction relative to a fixed notion of possibility, so that genuine conceivability is supposed to imply possibility. There are of course other ways to assign a content to this distinction. For example, we may employ the Cartesian notion of "clear" and "distinct" understanding, and stipulate that S

is genuinely conceivable if one clearly and distinctly can imagine a scenario in which S is true, and that S is apparently conceivable if one prima facie can imagine a scenario in which S is true. In this sense, it is not necessary that genuine conceivability implies possibility, at least not if we take "S is possible" to mean that S is *strictly* possible. Nevertheless, suppose that we do outline the distinction between genuine and apparent conceivability so that genuine conceivability, by stipulation or assumption, implies the relevant form of possibility. In that case, we obtain the relation between these concepts outlined in our explication of Arnauld's objection to Descartes (see section 4.2.1 above): genuine conceivability implies possibility, but apparent conceivability does not always imply possibility.

At this point, the modal epistemologist faces some difficult choices. First, we want the conceivability thesis to be true, but not at any cost. For example, we do not want to include "S is possible" as a necessary condition for "S is conceivable" in our *analysis* of the concept of conceivability. There is also a danger in dissociating conceivability (period) from apparent (or *experienced*) conceivability. It may be argued that the following thesis, which we may call the *transparency thesis* (cf. Sorensen 1992: 40; Almog 2002a: 18),[2] is plausible:

(1) If *x* apparently conceives of S, then *x* conceives of S.

For example, if it appears to me that I conceive that water is not H_2O, is there not a sense in which I actually conceive that water is not H_2O? On this account, conceivability data are transparent to the conceiver.

However, from (1), one could argue for the thesis:

(2) If a statement S is apparently conceivable for a conceiver *x*, then S is conceivable for *x*.

Suppose that S is apparently conceivable for *x*. Then *x* has the capacity to apparently conceive of S. But then, according to (1), *x* has the capacity to conceive of S. Thus, S is conceivable for *x*. According to this line of reasoning, a statement S is conceivable for a conceiver if it appears to the conceiver that he or she can conceive of S.

However, it has proven notoriously difficult to provide an account of conceivability according to which (i) conceivability data are transparent to the conceiver, and (ii) conceivability implies possibility. These problems will be dealt with at length in the present chapter.

---

[2] Sorensen notes that the thesis "if I believe that I am imagining *p*, then I am imagining *p*" has strong intuitive appeal.

When it comes to the distinction between absolute and relative conceivability, the distinctions that pertain to the manner in which something is conceived, and the distinction between apparent and genuine conceivability, there is, in each case, a short list of alternative explications. However, when we turn to substantial notions of conceivability, alternative explications abound. Gendler and Hawthorne (2002: 7) provide many possible senses of "to conceive of S":

> Rationally intuiting that it is possible that S.
> Realizing that ¬S is not necessary (i.e., realizing that S is possible).
> Imagining (that) S.
> Conjecturing that S.
> Accepting that S for the sake of argument.
> Describing to oneself a scenario where S obtains.
> Telling oneself a coherent story in which S obtains.
> Pretending that S.
> Make-believing that S.
> Supposing (that) S.
> Understanding the statement S.
> Entertaining (that) S.
> Mentally simulating S's obtaining.

Many of these items are reminiscent of what different philosophers have proposed that "S is conceivable" means. For example, "Pretending that S" and "Make-believing that S" are reminiscent of Ryle's analysis of what, roughly, "imagining S" means (see Ryle 1984 [1949]).[3] "Understanding the statement S" is reminiscent of one of Reid's proposals on how to interpret "*x* conceives of S." Reid offers two alternative interpretations of what it means to conceive of a proposition (Reid 1994: 377–8):

(3) To conceive of a proposition is to understand its meaning distinctly.

(4) To conceive of a proposition is to give some degree of assent to it, however small.

---

[3] According to Kind's interpretation of Ryle (Kind 1997), one objective of Ryle's behaviorist program is to show that even if imagining can be said to be a kind of "seeing mental images with the mind's eye," the mental images seen do not, thereby, exist. Ryle uses an analogy to a murder in a play. It would be correct to say that the actor (who plays the murderer) *pretends* to murder someone, but not to say that the actor *actually* murders someone (who pretends to be a real victim), thereby implying the existence of someone who is the object of the fake murder. According to Ryle, imagination can be explained along similar lines. Someone who imagines a murder does not, by the mind's eye, "see" a pretended murder; rather, this someone pretends being someone witnessing a murder. For details, see Kind 1997: 36–7.

There are also more complex notions of conceivability. Yablo (1993) discusses the following notions (of which he takes none to be the relevant *philosophical* notion of conceivability that he is looking for). Where 'S is believable for *x*' roughly means that S is compatible with everything *x* believes, Yablo suggests that "S is conceivable" can be taken to mean that:

> It is believable that S.
> It is believable that *possibly* S.
> One can imagine justifiably believing that S.
> One can imagine believing S truly.
> One can imagine believing something true with one's actual S-thought.

The problem with the first proposal on Yablo's list is that if I *know* that ¬S, then S will not be conceivable for me, since S will be unbelievable.[4] Therefore, since I know that I am sitting down, my standing up will not be conceivable for me. Nevertheless, on any reasonable understanding of what it means for something to be *possible*, I could have been standing up at the present moment. This problem is solved by the second proposal. The fact that I know that ¬S ("There is no tiger in the hallway") does not exclude the believability of the *possible truth* of the statement "There is a tiger in the hallway." In other words, I know that there is no tiger in the hallway, but I believe that there *could* have been a tiger in the hallway (under extraordinary circumstances). However, the second proposal (as well as the first) can be rejected on other grounds: consider a conceiver with nearly no information at all. Since most propositions will be compatible with what he or she knows, most propositions will be conceivable for him or her, including clearly impossible propositions (see our discussion in section 5.5 concerning (15) below). The third proposal is problematic since it is not clear how the "evidence" that justifies you in believing S should be imagined. For example, suppose that S is a mathematical theorem, and that a proof P for S can be imagined in *actually* convincing detail. Yablo notes that this seems to provide reason to believe that S is *true* rather than possible (1993: 32). The problems pertaining to the last two proposals are related to Kripke's thesis that there are a posteriori necessities (and hence, impossibilities that can

[4] We are assuming that if one knows that ¬S, then one believes ¬S, and S is not believable, that is, S is unbelievable. Of course, there are senses of 'believable' in which the believability of S is not blocked by my knowledge that ¬S. For example, suppose that S is "There is a tiger in the hallway," and further suppose that I know that there is no tiger in the hallway. Even though I know that there is no tiger in the hallway, I could still claim that it is believable that there is a tiger in the hallway, because the statement "There is a tiger in the hallway" does not, *as such*, involve any logical or conceptual inconsistency.

only be known a posteriori). Even if we will not address these particular proposals directly, we shall discuss the problems concerning such modal statements at length.

5.2 Absolute and relative notions of conceivability

One fundamental question pertaining to the notion of conceivability concerns what type of property *conceivability* is, or, equivalently, what type of predicate 'conceivable' is. Should we take "conceivable" to be a name of a *property* that statements possess, or should we take "conceivable" to stand for a *relation* that may or may not obtain between a conceiver and a statement? There seem to be at least three alternatives:

(i) First, we can understand conceivability to be an absolute, non-relational, property of statements. On this understanding, the statements that are conceivable are conceivable pure and simple, and they are so independently of the conceptual powers of each conceiver. (For each statement, it holds that it possesses the property of conceivability or it does not; the distribution of this property over the set of statements is absolute.)

(ii) Secondly, one can take conceivability to be a property of statements, although one (partially) defines what it means for a statement to possess this property in terms of knowledge, belief, and so forth—notions that imply the existence—or possible existence—of a conceiver. In this case, we take conceivability to be a *relational* (but not *relative*) property. This is arguably the case with "apriority." Apriority can perhaps be taken to be a genuine property of statements, although what it means for a statement to be a priori is normally defined in terms of possible *knowledge*, thus assuming, in a very ideal sense, the (possible) existence of a knowing subject.

(iii) Thirdly, one can take conceivability to be a relation that may obtain between a conceiver and a statement, such that the presence of certain cognitive or mental properties of the conceiver are crucially relevant for the relation to obtain. On this understanding, for any statement S to be conceivable, there must be a conceiver for which S is conceivable. Compare, for example, the relation "kicking." There must be someone who does the kicking (the "kicking agent") and something kickable that is kicked (a chair, for example).[5] Given a similar understanding of what "conceivable" stands for, conceivability will be a relative

[5] The "kicking" analogy appears in Kind 1997, where Kind compares "imagining" and "kicking." Almog (2002a) compares 'conceive' with activity verbs such as 'touch' and 'see'.

concept: a statement may be conceivable for a conceiver *a*, but not conceivable for *b*.

What are the virtues and vices of each alternative? We shall consider this question while giving examples of definitions from the literature.

5.2.1 Conceivability as a property of statements

Balog (1999) discusses a notion of conceivability according to which, at least on one possible understanding, conceivability is an absolute and non-relational property of statements. Balog's definition is the following (1999: 489):

(5) A statement S is conceivable if it is consistent with the totality of conceptual truths, that is, if ¬S is not a conceptual truth.

At least on some notions of conceptual truth—for example, if conceptual truth is understood in terms of objective relations between mind-independent abstract concepts—the notion of conceivability outlined is objective and mind-independent, and describes conceivability as an absolute property of statements. Given this understanding of the notion of conceptual truth, (5) promotes the idea of 'conceivable' as a statement predicate 'CONCEIVABLE'. Where *C* is the set of all conceptual truths, (5) can be interpreted:

(6) $\forall S(\text{CONCEIVABLE}(S) \leftrightarrow (\{S\} \cup C \nvdash \bot))$.

There is a clear sense in which conceivability, on the present account, is an absolute property of statements. What is the virtue of such an account? The foremost virtue, as I see it, is that skeptical remarks concerning our conceptual capacities—such as those in Arnauld's objection to Descartes' mind-body argument—become irrelevant. Arnauld's objection to Descartes' argument relies on the distinction between apparent and genuine conceivability (see section 4.2.1). However, the distinction between apparent and genuine conceivability, as we have outlined it in connection with Arnauld's objection, is only meaningful as long as the fact that conceivers may entertain inadequate conceptions is relevant. Given that the status of our conceptual capacities (or the conceptual capacities of *anyone*, for that matter) is irrelevant to the question of whether something is conceivable or not, it is also irrelevant to the question of whether the conceivability thesis is true. Thus, on the present account of conceivability, Arnauld's objection is irrelevant.

As for the vices of Balog's interpretation, I shall discuss them later, after having presented the other alternatives.

5.2.2 Conceivability as a relational property of statements

We shall now consider the position according to which conceivability is a relational property of statements, i.e., a property the obtaining of which implies the existence—or possible existence—of a conceiver. One such notion of conceivability is Chalmers' notion of *ideal* conceivability (Chalmers 2002a). Chalmers' notion of ideal conceivability can be spelled out in different ways. Here, however, I shall focus on the notion of ideal conceivability as presented in Chalmers 2002a. On this notion, a statement S is ideally conceivable when S is conceivable on ideal rational reflection (2002a: 147). How is this notion of conceivability to be understood? Chalmers considers different possible explications of the concept *ideal rational reflection*. Instead of postulating the existence of a (merely possible) ideal conceiver and define ideal rational reflection in terms of the mental states of that conceiver—a solution which naturally comes to mind—Chalmers defines this reflection (and thus ideal conceivability) in terms of "better" or "undefeatable" reasoning. Given this notion, we can say that

(7) A statement S is ideally conceivable iff there is a possible conceiver *x* such that S is prima facie conceivable for *x* with justification that is undefeatable by better reasoning.

The concept of "prima facie conceivability" is further explained as follows (Chalmers 2002a: 147):

(8) A statement S is prima facie conceivable for a conceiver *x* when S is conceivable for *x* on first appearances. That is, after some consideration, *x* finds that S passes the tests that are critical for conceivability.

We noted earlier that the concept of apriority plausibly can be described as a relational concept in this sense. Chalmers notes that there is "a fairly direct parallel between the idealization present in the notion of ideal conceivability and that present in the familiar notion of apriority" (2002a: 148). He continues:

> If I cannot know that P independent of experience, but another less limited being could do so, then it is a priori that P. And if I believe that P, but the justification for my belief is defeatable by better reasoning, then it is not a priori that P (unless there is another undefeatable justification). So the notion of apriority idealizes away from cognitive limitations in much the same way as the notion of ideal conceivability. This is not to say that either of these idealizations are perfectly clear, but at least the idealization [present in the notion of ideal conceivability] is a familiar one. (2002a: 148)

We shall return to these claims later. For now, one may note that Chalmers' attempt to define ideal conceivability in terms of undefeatability by better reasoning amounts to a less severe idealization than the attempt to define it in terms of the mental states of a *perfect* reasoner.

It is evident that a full analysis of the notion of ideal conceivability must be based on a full analysis of the notion of prima facie conceivability, which in turn amounts to a specification of "the tests that are critical for conceivability." As Chalmers notes, the specification will depend on a *substantive* notion of conceivability together with a specification of the manner in which the intended objects of conception should be conceived. One substantial notion of conceivability discussed by Chalmers (2002a: 147) is the following (cf. Balog's (5) above):

(9) A statement S is conceivable if S does not involve any apparent or explicable contradiction.

An agent-relative version of (9) is the following:[6]

(10) A statement S is conceivable for a conceiver *x* iff *x* cannot detect any contradiction in S.

Suppose now that we use (10) as a "template" notion of conceivability. What will it then mean that a statement S is *ideally* conceivable? First, assume that S does not involve any contradiction. Then, whenever S is *prima facie* conceivable *x*, S is also *ideally* conceivable, since the prima facie conceivability is undefeatable by better reasoning. Now, suppose to the contrary that S does involve a contradiction, but that some conceiver *a* finds S conceivable anyway. Then, prima facie, it seems as if things can turn out in two ways with respect to S. In the first case, the contradiction involved in S is not detectable even on ideal rational reflection, and *a*'s judgment is therefore undefeatable by better reasoning. In the second case, the contradiction is in principle detectable by better reasoning, and *a*'s judgment will therefore be defeated by some possible conceiver—this might be *a* itself—who finds S inconceivable. The possibility of the first case scenario is dependent on what type of idealizations we allow with respect to merely possibly conceivers. In acknowledging the possibility of first case scenarios, as we have done here, we have implicitly put a limit to idealizations with respect to possible conceivers. One could also consider a model where for each possible conceiver, there is a less limited conceiver. In

[6] Cf. Chalmers 2002a: 147. Chalmers takes (10) to express a variant of prima facie conceivability (where (8) above is applied to the substantial notion of conceivability outlined in (9)).

chapter 7, we shall turn to a discussion where we do not put any limit to idealizations with respect to possible conceivers.

5.2.3 Conceivability as a relation

As I have indicated in previous chapters, the origins of the conceivability thesis can be traced at least as far back as to Descartes' *Meditations*. In his writings, Descartes presents an account of conceivability as a *relation* that may obtain between a conceiver and an object or a state of affairs (involving objects possessing or lacking certain properties). Descartes' account of conceivability is presented mainly in terms of what must obtain with respect to the cognitive state of a conceiver in order for conceivability to obtain. However, his account can also be understood as involving restrictions on the object or state of affairs conceived: it must not be too complex. Since Descartes' account essentially involves reference both to a conceiver and a conceived entity, I shall take "conceivable," as understood by Descartes, to be a name of a relation "S is conceivable for *x*" where (i) S is the entity conceived (we shall continue to speak of statements), (ii) *x* is a conceiver, and (iii) where the specification of properties of *both* S and *x* are necessary in order to fully specify what "conceivable" means.

In his writings, Descartes carefully presents restrictions on *x* and S in "S is conceivable for *x*" which he takes to be *sufficient* for conceivability to imply possibility. These criteria were presented in detail in chapter 4, but I shall provide a brief summary of them here, and also expand a little on the discussion in that chapter. First, in order for "S is conceivable for *x*" to obtain with respect to a conceiver *x* and a statement S, *x* must have a clear and distinct conception of a situation in which S is true. A second criterion can be understood as pertaining both to the limits of a conceiver's conception, as well as to the nature or complexity of the entity conceived. According to Descartes, our ability to form clear and distinct conceptions is conditional. As he notes in Meditation 4, we sometimes *want* to make judgments about things that we do not have clear and distinct conceptions of. However,

> If […] I simply refrain from making a judgment in cases where I do not perceive the truth with sufficient clarity and distinctness, then it is clear that I am behaving correctly and avoiding error. But if in such cases I either affirm or deny, then I am not using my free will correctly. If I go for the alternative which is false, then obviously I shall be in error: if I take the other side, then it is by pure chance that I arrive at the truth, and I shall still be at fault since it is clear by the natural light that the perception of the intellect should always precede the determination of the will. (CSM 2, Med 4: 41)

Descartes argues that it is *clear by the natural light* that the perception of the intellect (which can be clear and distinct or unclear and undetailed) should always precede the determination of the will (on which one may want to make a judgment in any case). In other words, our will to affirm or deny must be restricted by an antecedent judgment as to whether the conception is clear and distinct. It is apparent that Descartes thinks that we can autonomously judge whether a particular conception is clear and distinct, and hence all that is needed of the conceiver in order not to make errors is to make the antecedent judgment whether the relevant conception *is* clear and distinct.

This second criterion can also, in a sense, be understood as a criterion pertaining to the thing conceived, and not only to the conceiver. Descartes writes:

> […] I have no cause for complaint on the grounds that the power of understanding or the natural light which God gave me is no greater than it is; for it is the nature of a finite intellect to lack understanding of many things […]. (CSM 2, Med 4: 42)

But, on the other hand:

> In order [for a created and thus finite intellect] to have adequate knowledge of a thing all that is required is that the power of knowing possessed by the intellect is adequate for the thing in question, and this can easily occur. (CSM 2, Rep 4: 155)

In other words, one may interpret the second criterion as an imperative that pertains solely to the relation between the conception and the will of the conceiver, but you can also interpret it as concerning the relation between the conceptual powers of the conceiver and *the thing conceived*, in so far as these two must match each other with respect to the powers of the first and the complexity of the second. (I do not, however, mean to say that these interpretations are incompatible.)

The third criterion—Descartes' requirement that conceptions must be complete—can be explained in terms of the first two, and since I have discussed this requirement at length in chapter 4, I shall not go further into it here. Here I shall just point out that Descartes' completeness criterion—as set forth in opposition to Arnauld's *adequacy* criterion—can be seen as an attempt to reestablish a notion of conceivability on which the question whether something is conceivable or not can be autonomously assessed by the conceiver. (Thus, on the assumption that clear and distinct conceivability implies possibility, the conceiver can autonomously assess whether something is *possible* or not.) As Almog (2002a: 18–19) points out, on Descartes' account, there is thus no *gap* between apparent conceivability and genuine conceivability: there is no gap

between seeming to conceive of something and succeeding in conceiving of the same thing. In other words, Descartes entertains (1) above, the thesis that if it appears to a conceiver *x* that *x* conceives of S, then *x* conceives of S.

In contemporary philosophy, Descartes' thesis that genuine conceivability, and thus possibility, can autonomously be established by a limited or "created" conceiver is lively debated and criticized, mostly in connection with worries such as Arnauld's (see chapter 4). Many authors are prepared to accept that some form of external conceivability—that is conceivability conditioned by facts external to the mental state of the conceiver—may imply possibility. For example, Chalmers (2002a) holds that certain forms of ideal conceivability imply possibility. However, it has become obvious in the debate that the conceivability thesis and Descartes' view that apparent conceivability implies genuine conceivability (or conceivability *simpliciter*) are difficult to combine. Thus, one of the foremost problems with relative notions of conceivability, on which conceivability is partly defined in terms of conditions on conceivers, is that no matter how clear and distinct a normal limited conceiver takes his conceptions to be, they may still be erroneous. This problem has led some authors, including Chalmers (2002a), to turn to notions of conceivability on which conceivability is defined in terms of ideal reflection. Whether this strategy is a viable one is a question we shall raise below. At this point, we note that it is only in connection with relative notions of conceivability—where conceivability is described as a relation between a statement and a conceiver—that the distinction between apparent and genuine conceivability is meaningful.

There are several contemporary accounts that, like Descartes', describe conceivability as a relation between a conceiver and a statement. Perhaps the best-known contemporary account of conceivability is presented by Yablo (1993).[7] On this account, conceivability is clearly a relative notion.

(11) A statement S is conceivable for a conceiver *x* iff *x* can imagine a possible world which he takes to verify S.

In considering Yablo's definitions, we shall take it for granted that the agent can imagine some worlds. Otherwise, no statement would be conceivable for the agent. Notice that Yablo describes conceivability as a relation between a conceiver and a statement that requires two things: the conceiver must be able to *imagine* a world that he or she *takes* to verify S. On the face of it, Yablo's

[7] Here, I shall speak of *statements* where Yablo speaks of *propositions*. This decision does not have any bearing on the content of the distinctions that we are about to discuss, nor on the fact that Yablo describes conceivability as a relation.

account is thus set forth in terms of both imagination and belief, which clearly requires the presence of a conceiver. The fact that Yablo requires a certain epistemic attitude towards the imagined world—that of *taking* the world to be a world that verifies S—also enables him to make finer distinctions between various ways in which a conceiver may *fail* to imagine a world than would otherwise be possible. For example, his distinction between *inconceivability* and *non-conceivability* (1993: 29–30) would not have been possible to make without reference to a conceiver:

(12) A statement S is *inconceivable* for a conceiver *x* iff for every world that *x* can imagine, *x* takes that world to falsify S.

Further:

(13) A statement S is *not conceivable* for a conceiver *x* iff for every world that *x* can imagine, *x* does not take that world to verify S.

(12) and (13) are not entirely clear. However, they can to some extent be clarified by pointing out that a world *w verifies* S iff S is true at *w*, and a world *w* falsifies S iff ¬S is true at *w*. Thus, a world *w verifies* a statement S iff the following holds: if *w* would have been the actual world, S would have been true, and a world *w falsifies* a statement S iff the following holds: if *w* would have been the actual world, ¬S would have been true. In terms of these distinctions, we can say that a statement S is *inconceivable* for a conceiver *x* iff for each world that *x* can imagine, *x* believes that it is such that if it had been actual, ¬S would have been true. Furthermore, a statement S is *non-conceivable* for a conceiver *x* iff *x* cannot imagine a world that *x* believes to be a world such that, had it been actual, S would have been true. If we assume that the conceiver is consistent, it is obvious that if a statement S is inconceivable for a conceiver *x*, then it is also not conceivable: a statement is inconceivable iff for every world *x* can imagine, *x* takes that world to falsify S, or, to *verify* ¬S. Thus, if S is inconceivable for *x*, there are no worlds *x* can imagine that *x* takes to verify S (since *x* takes all those worlds to verify ¬S).

5.3 Distinctions concerning the way in which something is conceived

In this section, I shall consider various distinctions concerning *the way* in which something is conceived.[8] These distinctions sometimes involve substantial notions of conceivability, and as before, we shall proceed by giving examples from the literature.

---

[8] These distinctions are all presented in Chalmers 2002a.

5.3.1 Positive versus negative notions of conceivability

First, one can distinguish between *positive* and *negative* interpretations of what it means that something is conceivable. An interpretation is of the *positive* variety if it requires that the conceiver *succeeds* in performing some mental or cognitive act. For example, Tidman (1994: 298–300) states that:

(14) A state of affairs is conceivable for a conceiver *x* iff *x* can form a mental image of the state of affairs in question.[9]

This interpretation of conceivability is clearly of the *positive* variety, since it requires that the conceiver *succeeds* in forming the relevant mental image. An interpretation is of the *negative* variety if it requires the *absence* of something on the conceiver's part. For example, the following interpretation has been attributed to William Kneale (Yablo 1993: 20):

(15) A proposition *p* is conceivable for me iff I have in mind no information which formally excludes *p*. (If I discover that ¬*p*, I withdraw my statement, because I then have information which entitles me to say that ¬*p*.)[10]

Another definition of conceivability in terms of information, but which is of the *positive* variety is due to Putnam (1990):

(16) A statement S is conceivable for a conceiver *x* iff *x* can imagine acquiring evidence that would justify him/her in believing S.

Kneale's interpretation is similar to Putnam's—although the former is of the negative variety and the latter of the positive—in that they both define conceivability in terms of information and belief. However, Kneale's definition explicitly describes conceivability as a "dynamic" notion: something may be conceivable for a conceiver, but may cease to be so at a later point.

To give further examples of how the distinction between positive and negative conceivability is used, it is clear that Yablo's notions of *inconceivability* and *non-conceivability* in (12) and (13) above employ this distinction. For a statement to be inconceivable for a conceiver, he or she must actually *succeed* in imagining some ¬S-worlds. On the other hand, for S to be non-conceivable for the same conceiver, it is sufficient that the conceiver *fails* to imagine any S-worlds.

---

[9] Tidman proposes (14) as an interpretation of Hume's account of conceivability in Hume 1978 [1739].

[10] I take it that I have no "information in mind" which formally excludes *p* iff *p* is consistent with the set of all my beliefs.

The (explicit) distinction between positive and negative forms of conceivability possibly originates with Van Cleve (1983). Van Cleve argues that there is a certain kind of intellectual perception, which allows us to recognize, for example, that a certain mathematical proposition is true or that a certain proposition or statement is possible. In order to distinguish between ordinary perception and the kind of intellectual perception that Van Cleve has in mind, we may call the latter type of perception "intuitive" perception. In his paper, Van Cleve does not actually use the terms 'positive' and 'negative'. However, he proposes definitions of 'strong' and 'weak' conceivability along the following lines (Van Cleve 1983: 37; see also Tidman 1994: 307):

(17) A proposition *p* is *strongly conceivable* for a conceiver *x* iff *x* intuitively sees that *p* is possible.

(18) A proposition *p* is *weakly conceivable* for a conceiver *x* iff *x* does not intuitively see that *p* is impossible.

Here, it is obvious that for strong conceivability to obtain, the conceiver must succeed in doing something: "seeing" that *p* is possible. By contrast, for weak conceivability to obtain, only the absence of something is required, namely the conceiver's ability to see that *p* is impossible.

The clearest treatment of the distinction between positive and negative notions of conceivability can be found in Chalmers 2002a. What is particular to Chalmers' account is that he associates positive and negative conceivability, respectively, with different kinds of cognitive and mental activities. With respect to positive notions of conceivability, Chalmers reasons as follows:

> One can place the varieties of positive conceivability under the broad rubric of *imagination*: to positively conceive of something is to imagine (in some sense) a specific configuration of objects and properties. It is common to imagine situations in considerable detail, and this imagination is often accompanied by interpretation and reasoning. When one imagines a situation and reasons about it, the situation is often revealed as a situation in which S is the case, for some S. When this is so, we can say that the imagined situation *verifies* S, and that one has *imagined that* S. (Chalmers 2002a: 150)

Chalmers thus defines positive conceivability as follows:

(19) S is *positively conceivable* for a conceiver *x* when *x* can imagine a situation that verifies S.

Chalmers notes that (19) is indebted to Yablo (1993). Yablo's definition of conceivability is given in (11) above, according to which a statement S is conceivable for a conceiver iff he or she can imagine a possible world which he

or she takes to verify S. Note, though, that one component of Yablo's definition (11) seems to be missing in Chalmers' (19). According to Yablo's definition, a statement S is conceivable if the conceiver *takes it that* the imagined world (or situation) verifies S. According to Chalmers' definition, a statement S is conceivable if the imagined situation (or world) verifies S *period*. In other words: on the face of it, it seems as if Chalmers' definition takes conceivability to be matter of whether an imagined situation in fact verifies a statement S, not a matter of whether the conceiver *believes* that the imagined situation verifies S. However, from the passage cited, it seems that Chalmers takes the conceiver's reasoning about what is imagined to be crucial to conceivability.

Chalmers further holds that the negative variety of conceivability can be associated with notions of conceivability according to which a statement is conceivable if S is not *ruled out* (2002a: 149). For example, Kneale's definition (15) says that a statement S is conceivable for x if S is not ruled out by what x knows. Chalmers notes that this notion of conceivability seems to be in conflict with a notion of conceivability, according to which a statement S can be conceivable even if one knows that ¬S. However, this problem can be solved by specifying how S is to be "ruled out" in order to be inconceivable.

5.3.2 The distinction between primary and secondary conceivability

In section 4.4.2 above, we explained the distinction between primary possibility and secondary possibility. There we said that a statement S is *possible with respect to its primary intension*, or *primarily possible* for short, if the following obtains:

(20) There is at least one logically possible world w such that the proposition $[\![\mathrm{S}]\!]_w$ is true in w.

That is, S is primarily possible iff there is at least one logically possible world w such that the statement S, when uttered in w, is true in w. We can further identify the *primary proposition* associated with S as the set S_1 of worlds w such that S, when uttered in w, is true in w

(21) $S_1 = \{w : [\![\mathrm{S}]\!]_w \text{ is true in } w\}$.

As an example of a primary possibility, we gave the following statement:

(22) Water $\neq$ H_2O.

(22) is primarily possible since there are worlds w such that the proposition $[\![\text{Water} \neq H_2O]\!]_w$ (that is, the proposition expressed by (22) in w) is true in w (one such world is Twin World). We also said that a statement S is *possible*

with respect to its secondary intension, or *secondarily possible* for short, if the following obtains:

(23) There is at least one logically possible world w such that the proposition $[\![S]\!]_@$ (where '@' is a name of the actual world) is true in w.

We can identify the *secondary proposition* associated with S with the set S_2 of worlds w such that $[\![S]\!]_@$ is true in w:

(24) $S_2 = \{w : [\![S]\!]_@ \text{ is true in } w\}$.

Consider (22) again. The proposition $[\![\text{Water} \neq H_2O]\!]_@$ (that is, the proposition expressed by (22) in the *actual world*, or, the proposition that (22) *actually expresses*) is a *necessarily false* proposition, that is, a proposition that is false in all logically possible worlds, namely that $H_2O \neq H_2O$.[11] In other words,

$\{w : [\![\text{Water} \neq H_2O]\!]_@ \text{ is true in } w\} = \varnothing$.

Once we have distinguished between the two propositions associated with each statement S, we can also in a straightforward way define what it means that a statement is *primarily* or *secondarily* conceivable:

(25) A statement S is *primarily conceivable* for a conceiver x iff x can imagine a world w in which the proposition $[\![S]\!]_w$ is true.

(26) A statement S is *secondarily conceivable* for a conceiver x iff x can imagine a world w in which the proposition $[\![S]\!]_@$ is true.

In summary, the distinction between primary and secondary conceivability can be described as a distinction pertaining to *the proposition conceived of*. In the one case, you conceive of one proposition being true, and in the other case, you conceive of another proposition being true. However, this explanation is not satisfactory. We want to understand what it means, in cognitive terms, to conceive of the first proposition rather than the second. Fortunately, Chalmers explains what primary versus secondary conceivability comes to in cognitive terms. In short, we can say that primary conceivability involves a *counteractual* assumption, and secondary conceivability a *counterfactual* assumption. Consider (25) and (26) in relation to (22). (25) says that (22) is conceivable for a conceiver x iff x can imagine a world such that *the proposition expressed in that world* by (22) is true in that world. For example, in Twin World, (22) expresses the proposition that $XYZ \neq H_2O$—a proposition that, according to

[11] For details, see section 4.4.2.

Kripkean principles of identity, is true in *all* possible worlds. Therefore, all the conceiver has to do in order for (22) to be primarily conceivable is to imagine Twin World, a world *in* which (22) expresses a true proposition.

Using cognitive terms, the above can be simplified. Given that our world is the actual world, (22) does *in fact* express the proposition that $H_2O \neq H_2O$. However, to determine whether a statement is primarily conceivable or not, we must, in the typical case, disregard what proposition the statement *in fact* expresses. There are different ways of describing how to do this. First, we can consider what is conceivable for us on the assumption that everything we have learned about the empirical world—that is, all our a posteriori knowledge—turns out to be mistaken (or, at least, that it *may be* mistaken). On this assumption, the question whether a statement is primarily conceivable amounts to the question if the world might actually be such that S is true. For example, on the assumption that everything we have learned about the actual world turns out to be mistaken, (22) is conceivable: water might be H_2O, XYZ, or even an element. Secondly, but equivalently, one might turn directly to the part of our knowledge that is left when all our a posteriori knowledge has been (hypothetically) suspended, namely our a priori knowledge. In order to find out whether S is primarily conceivable we can ask: might S be true for all we know a priori? If the answer is yes, then S is conceivable in the sense intended. (For example, it might be true for all I know a priori that water is XYZ.)

I now turn to the notion of secondary conceivability. In section 2.5, we initiated a discussion that we shall return to in chapter 6. We noted that van Inwagen (1998) and Weigel (2000) both argue that a posteriori modal knowledge stems from what can be broadly described as a priori principles of modality and empirical knowledge. What could such a priori principles of modality be? I shall argue that at least the logical laws that govern the concept of identity and logical principles of individuation are such principles. (These ideas will be presented in detail in chapter 6.) These laws and principles include *at least* the following theses: that all things are necessarily self-identical:

(27) $\forall x \, \Box(x = x)$,

Leibniz' Law:

(28) $\forall x \forall y (x = y \rightarrow (Fx \rightarrow Fy))$,

that all identical things are necessarily identical (which follows from (27) and (28)):

(29) $\forall x \forall y (x = y \rightarrow \Box(x = y))$,

and that (29) is true in all possible worlds:

(30) $\Box\forall x\forall y(x = y \rightarrow \Box(x = y))$.

(27)–(30), I argue, are all a priori true theses pertaining to the concept of identity. We know a priori that each individual thing is necessarily self-identical, we know a priori that any two identical things must have the same properties, and so on. Many philosophical problems are grounded in the fact that we do *not* know a priori when the antecedents in (28)–(30) are true. For example, in Kripke's famous puzzle about Pierre's beliefs about London (Kripke 1979), Pierre does not know that London = Londres, and thinks that Londres is pretty but London is not. However, Pierre's beliefs are in apparent conflict with (28), according to which the propositions that London = Londres and that Londres is pretty logically implies that London is pretty. However, that London = Londres and that London is ugly can only be known a posteriori.

These remarks bring us closer to a cognitive characterization of secondary conceivability. Whereas any attempt to determine whether a statement is primarily conceivable or not involves the assumption that everything we know about the actual world is (or can turn out to be) false, the attempt to determine whether a statement is secondarily conceivable or not involves the diametrically opposed assumption. Here, facts about the actual world are considered as *given* prior to the question about conceivability. At least, this is how nearly all examples in *Naming and Necessity* (1980) are presented. Kripke asks whether cats could have been demons *given* that they are animals (1980: 125–6), and whether, *given* that gold has the atomic number 79, it could be gold without having this atomic number (1980: 123–4). In this sense, (22) above is *inconceivable*: *given* that water = H_2O, it is inconceivable that water ≠ H_2O. In summary then, secondary conceivability can be characterized in cognitive terms as follows: a statement S is secondarily conceivable iff it is conceivable that facts could have been the way that S *actually* describes them to be.

5.3.3 The distinction between ideal and prima facie conceivability

Finally, there is the distinction between ideal and prima facie conceivability. I have already presented this distinction in section 5.2.2 above, where I argued that on the notion of ideal conceivability, conceivability can be described as a relational property of statements. However, this distinction can also be seen as a distinction pertaining to the manner in which something is conceived. While employing the same substantial notion of conceivability, a statement S can be said to be ideally conceivable iff S (i) is conceivable in the sense implied by the

substantial notion, and (ii) the conception fulfills certain requirements in terms of completeness, clearness, and so on. And S can be said to be prima facie conceivable iff S fulfills (i) and is, at least as far as the conceiver can judge by introspection, complete, clear, and so on. (This is just one of many possible proposals concerning how to provide the terms 'ideal' and 'prima facie' with content.)

With the distinctions between ideal and prima facie conceivability, positive and negative conceivability, and primary and secondary conceivability in place, I shall close my discussion of the manner in which something is conceived by considering two notions of conceivability which both involve each of these distinctions. Chalmers focuses on these particular dimensions of conceivability because he believes that on the right notion of conceivability, conceivability does indeed imply possibility (2002a: 171). One variant of the conceivability thesis that Chalmers takes to be valid is:

(31) Ideal primary conceivability entails primary possibility.

He further thinks that the following thesis is plausible:

(32) Ideal secondary conceivability entails secondary possibility.[12]

In terms of the distinctions discussed in sections 5.3.1–5.3.3, we can explicate (31) and (32). First, we consider (31). We shall arbitrarily focus on (9) above as our substantial interpretation of conceivability. Furthermore, Chalmers argues that positive conceivability is generally a better guide to possibility than negative possibility (2002a: 160). I shall just accept this claim, and focus on positive conceivability.[13] Now, if a statement is substantially conceivable iff a conceiver cannot detect any contradiction in the hypothesis that S, what does it take for the same statement to be *ideally primarily positively conceivable*? First, in order to ensure that the conception is of the positive variety, we adopt Chalmers' definition of positive conceivability in (19), according to which a

---

[12] Chalmers also believes that the following thesis is very likely to be true: *secunda facie primary positive conceivability is an extremely good guide to primary possibility*. It is not entirely clear how to understand what Chalmers means by *secunda facie* conceivability, but from the examples he provides, it seems as if something is *secunda facie* conceivable iff it is conceivable for a competent conceiver after in-depth considerations. However, since we have made the preliminary decision to take the conceivability thesis to assert that conceivability *suffices* for possibility (see section 3.3), we shall focus on (31) and (32).

[13] It may further be argued that positive conceivability entails negative conceivability. However, I shall not rely on this thesis. At any rate, it is clear that positive conceivability is a stronger notion than is negative conceivability.

statement S is *positively conceivable* for a conceiver *x* when *x* can imagine a situation that verifies S.

Secondly, we see that given that our substantial interpretation of conceivability is (9), we need not be concerned with the distinction between primary and secondary conceivability. This is because the distinction between primary and secondary conceivability collapses if the only desideratum on conceivability is logical consistency. In section 5.2.2, we noted that Chalmers thinks that a full analysis of ideal conceivability will depend on a substantive notion of conceivability. This also holds true of the distinction between primary and secondary conceivability: in relation to some substantive notions of conceivability, the distinction between primary and secondary conceivability will be superfluous, but in relation to other notions it will be crucial.

Finally, given that Chalmers associates positive conceivability with imaginability—see (19)—we can say that S is *ideally* positively conceivable for a conceiver *x* iff *x* can imagine a possible world that *x* takes to verify S, and *x*'s statement to this effect is undefeatable by better reasoning. A more complex notion of conceivability would result if we also took the primary/secondary distinction into account. However, what such a notion might look like will to some extent be clear from our explication of ideal *secondary* positive conceivability, to which I now turn.

(32) says that if a statement S is ideally secondarily positively conceivable, then it is also secondarily possible. (For the notion of secondary possibility, see (23) above.) In (26) above, we explicated the notion of secondary conceivability as follows: a statement S is *secondarily conceivable* for a conceiver *x* iff *x* can imagine a world *w* in which the proposition $[\![S]\!]_@$, that is, the proposition actually expressed by S is true. In the discussion that followed, we noted that this explication is not entirely satisfactory, and that we would like an explication in cognitive terms. The explication we proposed was the following: a statement S is secondarily conceivable iff it is conceivable that facts could have been the way that S actually describes them to be.

At this point, let us pause to consider the fact that this explication of secondary conceivability is still set out *in terms of external facts*. According to our explication, S is conceivable in the sense intended iff it is conceivable given *all the relevant facts* about the actual world. However, in any given case, how can a conceiver *know* that all the relevant facts about the actual world are taken into account? For a conceiver who does not possess the relevant information, statements such as "Water $\neq H_2O$," "Gold does not have the atomic number 79," and so forth, will *appear* secondarily conceivable, but given our

explication, they are not, because they will not be conceivable to a conceiver that has this information.

In other words, we seem to face the familiar gap between *apparent* and *genuine* conceivability. A statement S is *genuinely* secondarily conceivable iff S is conceivable given all relevant empirical information. Thus, whether a particular statement *is* genuinely secondarily conceivable or not is determined both by facts about the state of the conceiver (whether a certain cognitive state obtains, whether the conception is of the positive variety, and so forth) *and* by facts about external reality (the empirical facts in question). Usually, we do not antecedently expect limited conceivers to have de facto complete information about all relevant facts involved in making judgments about empirical states of affairs. (This is why one often feels the need to qualify statements about law-like relations with a *ceteris paribus* clause.) This is where Chalmers' notion of *ideal* secondary conceivability comes in. Faced with this problem, he goes on to define ideal positive secondary conceivability in terms of (positive) prima facie conceivability on *complete empirical knowledge*.

> One might say that a subject prima facie secondarily conceives of S when the subject imagines a situation and judges that if that situation had obtained, S would have been the case. One can say that S is ideally secondarily conceivable if S is prima facie secondarily conceivable, and if the secondary conceivability is not defeatable by idealized rational reflection and complete empirical knowledge. To avoid trivializing the link between conceivability and possibility here, it is probably best to restrict the empirical knowledge in question to nonmodal knowledge. (2002a: 159)

Much needs to be said about this proposal. However, I shall delay my comments until my general discussion at the end of this chapter. However, I want to point out that Chalmers' definition of *prima facie positive* conceivability, which in general terms comes to the following,

(33) A statement S is prima facie positively conceivable for a conceiver *x* iff *x* can imagine a situation and judges, after some consideration, that if that situation had obtained, S would have been the case,

is nearly equivalent to Yablo's definition of conceivability *simpliciter* (see (11) above). We also noted above that in Chalmers' definition of positive conceivability *simpliciter* ((19) above), the reference to the epistemic attitude of the conceiver—which seems central to Yablo's definition—is dropped altogether. Obviously, in Chalmers' definitions of *prima facie* versions of conceivability, such as (33), this reference reappears.

Given the notion of prima facie positive secondary conceivability outlined, Chalmers notion of ideal positive secondary conceivability can be explicated as follows. In giving this explication, I shall draw on discussions and definitions introduced in previous sections of this chapter. For example, since "*w* verifies S" (where *w* is a situation or a possible world), as used by Yablo, means that had *w* been actual, S would have been true (or "the case"), I shall for the sake of brevity replace the latter notion with the former. (The fact that these notions have the same meaning for Yablo further supports the thesis that Chalmers' (33) is equivalent to his (11).)

(34) A statement is ideally, positively, and secondarily conceivable for a conceiver *x* iff (i) *x* can imagine a situation that *x* takes to verify S, and (ii) there is no possible conceiver that can defeat *x*'s statement to this effect by means of better reasoning and/or more extensive empirical knowledge.

Again, there are prima facie two possibilities if the desiderata on conceptions in (34) are fulfilled. In the first case, what the conceiver *x* conceives is possible, and his/hers conception is (therefore) ideal. In the second case, what the conceiver *x* conceives is not possible, but the conception is still ideal, since it is undefeatable by better reasoning. However, as we noted in section 5.2.2, the possibility of the latter scenario, a scenario where a statement S is ideally conceivable without being possible, is dependent on what type of idealizations we allow with respect to possible conceivers.

The question whether Chalmers' conceivability theses (31) and (32) are plausible, and to some extent Chalmers' arguments for them, will be addressed in section 7.3.1 below.

5.4 Genuine and apparent conceivability

Left to discuss are the various substantial notions of conceivability and the distinction between apparent and genuine conceivability. When it comes to substantial notions of conceivability, I have already presented all of the notions that I want to discuss, as examples of the various distinctions made above. In section 5.5, I shall present counterexamples to the conceivability thesis when interpreted in terms of some of these notions of conceivability. I shall also present examples showing that some of these notions do not satisfy certain desiderata that we require a satisfactory analysis of conceivability to satisfy.

My discussion that follows will center on the distinction between apparent and genuine conceivability, which has been in focus throughout this essay. The

discussion up to now, as well as the discussion to follow, suggests that any attempt to evaluate the conceivability thesis will involve a dilemma set out in terms of this distinction. In order to address this dilemma, I shall present two general views on conceivability and possibility that seem to be incompatible. These views have already been hinted at in our discussion of Descartes and Arnauld in section 4.2, and in our discussion of Kripke and Chalmers in sections 4.4.1 and 4.4.2.

5.4.1 The misdescription model of modal error

As we have previously indicated, the origin of the distinction between genuine and apparent conceivability can be traced back to the discussion between Descartes and Arnauld. However, in Arnauld, it seems that this distinction is employed to do a dual job. On the one hand, genuine conceivability may be described as such because it is taken to imply possibility. In this sense, the distinction between genuine and apparent conceivability is in the first place a distinction that pertains to the relation between conceivability and possibility. On the other hand, Arnauld provides epistemological conditions on genuine conceivability, and thus he (also) provides a *content* to the notion of genuine conceivability. In this latter sense, apparent conceptions are just conceptions that fail to meet the demands on genuine conception.

In the contemporary debate, the distinction between apparent and genuine conceivability is often used in the first sense above, that is, the sense in which this distinction pertains to the relation between conceivability and possibility. However, contemporary philosophers are less inclined to provide the notion of genuine conceivability with a determinate content. For example, as we discussed in section 4.4.1 above, Kripke (1980) argues that we sometimes *misdescribe* the scenario we *actually* conceive of, to the effect that we conclude that another, epistemically indistinguishable scenario, is possible. In Kripke, there is no universally applicable method for avoiding such mistakes, and Kripke never considers, such as Arnauld does, imposing requirements on *conceptions* in order to avoid such mistakes.

Arnauld holds that one must know all the essential properties of the thing(s) involved in order for genuine conceivability to obtain. Kripke, on the other hand, seems to contest this requirement. At least, the following passage from *Naming and Necessity* can be read as an implicit criticism of Arnauld's position:

> [S]upposing Nixon is in fact a human being, it would seem that we cannot think of a possible counterfactual situation in which he was, say, an inanimate object; perhaps it is not even possible for him not to have been a human being. Then it

> will be a necessary fact that in all possible worlds where he exists at all, he is human or anyway he is not an inanimate object. This has nothing to do with any requirement that there be purely *qualitative* conditions for Nixonhood which we can spell out. And should there be? Maybe there is some argument that there should be, but we can consider these questions about *necessary* conditions without going into any question about *sufficient* conditions. Further, even if there were a purely qualitative set of necessary and sufficient conditions for being Nixon, the view I advocate would not demand that we find these conditions *before* we can ask whether Nixon might have won the election, nor does it demand that we restate the question in terms of such conditions. We can simply consider *Nixon* and ask what might have happened to *him* had various circumstances been different. So the two views, the two ways of looking at things, do seem to me to make a difference. (1980: 46–7)

A lot can be gained from this passage. As often, Kripke does not mention any particular philosophers or arguments. However, I shall take the liberty of filling in the blanks according to my own reading. Firstly, the two views mentioned at the end of the quote can be construed as Kripke's own view and Arnauld's view respectively. Secondly, Kripke says that on his view, necessary and sufficient conditions (for Nixonhood) need not be found in order for us to be able to make *de re* modal judgments about Nixon. This seems to be in opposition to Arnauld, according to whom all such judgments must be based on adequate knowledge, that is, knowledge of all essential (or necessary) properties of the thing(s) in question. Finally, Kripke's remark that "even if there were a purely qualitative set of necessary and sufficient conditions for being Nixon" can be understood as expressing the doubt whether there *is* such a set of conditions.

With these considerations (and those of section 4.4.1) in mind, it would be wrong to describe Arnauld and Kripke as employing the same notion of (genuine) conceivability. In what follows, I shall use 'genuine conceivability' to mean the notion of conceivability on which conceivability is taken to imply possibility, but I shall, unless otherwise stated, leave it open whether the *content* of this notion is to be understood à la Arnauld, à la Kripke (or à la Chalmers). What I shall take to be common to most philosophers who employ the distinction between apparent and genuine conceivability (regardless of the specific ways of endowing the distinction between genuine and apparent conceivability with content) is that they employ what I shall call the *misdescription model of modal error*. For example, both Kripke and Chalmers argue that we sometimes *misdescribe* what we actually conceive of as something else, in the sense that we infer the possibility of what we sought to conceive of, rather than the possibility of what we actually conceived of. To repeat the basics of the account provided in sections 4.4.1 and 4.4.2, an example of how

such a "misdescription" mistake can occur would be the following. Suppose that I aim to conceive that water is not H_2O, and that I actually conceive of some watery stuff being something else than H_2O. I thus infer that it is possible that water is not H_2O (a metaphysical impossibility, which was what we sought to conceive of) instead of inferring the possibility of what I actually conceived of, that is, of watery stuff being something else than H_2O (a metaphysical possibility).

The misdescription model of modal error can be applied, it seems, regardless of what mental capacities one relies on when defining what it means that something is conceivable. For example, some philosophers argue that if conceivability is interpreted in terms of pictorial representability—that is, roughly in terms of (14) above—impossible objects such as Escher staircases are conceivable in the relevant sense.[14] Now, given that one wants to save the conceivability thesis (on this interpretation of what 'conceivable' means), an alternative explanation can be provided by the misdescription model. You can argue that the *actual* object of the relevant representation has been misidentified, and that the object *actually* represented *is* possible. According to this explanation, it is plausible that the object of the representation has been *erroneously* identified as an Escher staircase (i.e., an impossible object). The object of the representation can also be identified as the Escher picture itself, which is indeed possible, since it actually exists.[15]

The misdescription model also applies if one takes conceivability to involve descriptional elements. Yablo (1993: 31–2) suggests that his imagining of a computer producing a proof unanimously hailed as the refutation of Goldbach's conjecture falls short of being an imagining of a scenario in which Goldbach's conjecture is false. Why? Because his imagining may equally well be described as an imagining of a scenario in which Goldbach's conjecture is *true*, and where the proof produced by the computer is invalid (although unanimously, but erroneously, accepted as valid).

It is clear that the misdescription model of modal error does not respect the authority of first-person judgments about the content of one's conceptions and imaginations. This has caused indignation.[16] Why should my Goldbach

---

[14] See Tidman 1994; Kind 1997.

[15] There are actually good arguments why the object of the relevant representation should be identified as the Escher picture itself rather than the impossible Escher staircase. The arguments pertain to the spatial properties of the object imagined. The represented object, *as represented*, seems to share the spatial properties of the Escher picture, not of the Escher staircase. I shall briefly return to these arguments in section 5.5.

[16] Cf. Sorensen 1992: 40.

imagination be described as an imagination of the latter scenario rather than the former? Analogously, if I say that I am imagining an Escher staircase—an impossible object—how can you say that I am mistaken about what I imagine? One way in which these objections can be met on the misdescription model is to make it plausible that neither of these two imaginations is carried out in enough detail to decide between the alternative interpretations. The misdescription theorist can then argue that as long as there is an alternative interpretation at hand, according to which what has been imagined is not something impossible but something *possible*, the mere existence of such an explanation invalidates the intended counterexample to the conceivability thesis. Apart from this, it is clear that the conceiver cannot simply *will* his conception to be of one state of affairs rather than another. Suppose that first-person judgments are to be given absolute authority. Then, I can simply *will* the representation

---

to be a representation of a squared circle (side-view).[17] Moreover, why should first-person judgments about the content of conceptions be given absolute authority? We routinely question first-person judgments grounded in evidence from perception: we ask whether people really saw what they think they saw, and if they really heard what they think they heard. We also present our own suggestions as to what it may have been that they *actually* saw, or heard. Why should evidence from conceivability be subjected to special treatment?

In summary, the views on conceivability and the conceivability thesis promoted by the misdescription model of modal error can be expressed by the following theses.[18] (In formulating these theses, we return to our former convention of speaking of statements.)

(35) If a statement S is genuinely conceivable, then S is possible.

The distinction between apparent and genuine conceivability can be summarized as follows:

(36) A statement S is *apparently conceivable* for a conceiver *x* if *x* believes or experiences that S is conceivable for *x*.

---

[17] Cf. Sorensen 2002: 343–4.

[18] The following theses need not match the beliefs of all philosophers that employ the misdescription model of modal error or something similar to it. Here, we abstract from differences between individual philosophers in order to demarcate a general position with respect to conceivability and the conceivability thesis which can be contrasted with other, equally general positions.

(37) A statement S is *genuinely conceivable* for a conceiver *x* if (i) S is apparently conceivable for *x* and (ii) *x* is de facto not confused as to what statement it is that *x* conceives of.

Here, epistemological or otherwise "qualitative" restrictions on conceivability in general, and genuine conceivability in particular, can be imposed (à la Arnauld, for example). Finally:

(38) Modal errors—counterexamples to the conceivability thesis—can be explained by means of the misdescription model, as outlined above.

5.4.2 The misdescription model and the conceivability thesis

It is easy to see how the conceivability thesis can be saved, come what may, by means of the misdescription model. The proponents of the misdescription model are Kripke and Arnauld. What it means, or, what is required, in order for a conceiver to have a genuine conception of a certain state of affairs according to Kripke is difficult to tell (although we have provided some suggestions above). On the other hand, we have provided a detailed account of what this means according to Arnauld: in order for something to be genuinely conceivable for a conceiver, the conceiver must have an *adequate* representation of the things involved in the representation. In order for the conceiver to have an adequate representation of these things, he or she must know each essential property of them. (If this criterion is fulfilled, we say that the conceiver possesses *adequate knowledge* of them.) Thus, for something to be genuinely conceivable on the misdescription model—at least when the notion of genuine conceivability is endowed with a content à la Arnauld—it must be nothing less than *possible*. To conceive of something *impossible*, on Arnauld's account, would be to conceive that, for example, *Fy*, where *F* is incompatible with the essence of *y*. But if you conceive of something that is incompatible with the essence of *y*, you cannot have a genuine conception of *y*, in which case you cannot have thus conceived.

At least according to Almog (2002a), the same seems to apply to Kripke's account. Almog writes:

> In many ways, [Kripke] is very close to Arnauld because of Kripke's fundamental distinction between seeming to conceive and really conceiving. Also, like Arnauld, Kripke emphasizes that […] we may turn out to be wrong about what we really conceive of; it all depends on whether we respected the essence of the target object, and this essence may not be transparent to us. (2002a: 43)

To this, I merely add that to "respect the essence of the target object" may, with respect to conceptual as well as relational content, very well mean different things for Arnauld and Kripke (as suggested above). Almog argues that on Kripke's account,

> water is not *really* imaginable without oxygen *because* it is not possible for water to exist without oxygen. Real imaginability about an object *x* is dependent on and posterior to what is really possible for *x*. (2002a: 50)

Thus, at least according to some philosophers that employ the misdescription model of modal error, possibility *determines* genuine conceivability. Given the thesis that nothing is genuinely conceivable unless it is possible, the conceivability thesis, interpreted in terms of genuine conceivability, becomes tautologous. One may also note that the conceivability thesis is tautologous on an account where *essence* is held to determine both genuine conceivability and possibility. Such accounts will be considered in the next chapter.

5.4.3 The misdescription model and the transparency thesis

(36) above asserts that a statement S is apparently conceivable for a conceiver *x* iff *x* believes or experiences that S is conceivable for *x*. Some philosophers think that this is what it means that a statement is conceivable *period*. At the beginning of this chapter, we noted that it may be argued that the *transparency thesis* (1), which states that if it appears to a conceiver that he or she conceives of S then he or she conceives of S, is intuitively plausible. (For example, if it appears to me that I can conceive that water is not H_2O, is there not a sense in which I actually can conceive that water is not H_2O?) Almog (2002a: 18) argues that the transparency thesis (1) constitutes Descartes' "starting point" in his mind-body argument and in his subsequent dialogue with Arnauld. As we noted above (section 5.2.3), Almog's point is that on Descartes' account, there is no *gap* between apparent conceivability and genuine conceivability—there is no gap between seeming to conceive of something and succeeding in conceiving of the same thing. However, one should bear in mind that even Descartes distinguishes between conceptions that are clear and distinct and those that are not.

Turning to contemporary philosophers, I believe that the ideas of Putnam (1990) support the thesis (1). Putnam begins his discussion pertaining to Kripke's *Naming and Necessity* by raising the following doubt:

> Although everyone—including me—read *Naming and Necessity* as denying the "conceivability implies possibility" inference, I am now not so sure we read it correctly. (1990: 55)

Putnam then proceeds to outline an interpretation of what it means that something is conceivable which, at first, appears to explain modal mistakes in terms of the misdescription model: what appears to be conceivable need not be genuinely conceivable.

But Putnam also argues in a different way. He asks us to consider a purported proof P of a mathematical theorem T, and to assume that P is in fact a proof of T. Putnam argues that if we make the minimal assumption that conceivability implies logical possibility—an assumption which we shall grant for the sake of the argument—then it is not conceivable that P is not a proof of T unless P is in fact not a proof of T. Suppose now that I claim to have conceived that P is not a proof of T, that is, that there is a mistake in P, and that I, via the conceivability thesis, thus conclude that it is possible that P is no proof of T. On the misdescription model, one would explain my mistaken possibility judgment with reference to a mistaken *conceivability* judgment. I did not actually conceive that P is no proof of T, the misdescription theorist would say. Rather, what I actually conceived of is that there is a mistake in (using Putnam's term) an *epistemic counterpart* of P. For example, suppose that P is Andrew Wiles' proof of Fermat's conjecture. To a mathematical novice, almost any arrangement of symbols P*—imagined or real—with roughly the same length and appearance as P would be an epistemic counterpart of P.

At this point, Putnam argues as follows:

> If an *epistemic counterpart* [P*] of this proof [that is, P]—a proof with respect to which I have the same evidence that it is a proof of this theorem [T]—could be wrong, then it is right to say "it is conceivable that there is a mistake in the proof [P]." (1990: 62)

This passage is crucial. It is evident that Putnam does not have the same understanding of what it means that something is conceivable (period) as the misdescription theorist. The misdescription theorist identifies conceivability (period) with the notion of genuine conceivability outlined in (37) above, and discards the notion of apparent conceivability outlined in (36) as, strictly speaking, irrelevant to the truth of the conceivability thesis. However, it is evident from Putnam's discussion that he does *not* take the notion of apparent conceivability outlined in (36) to explain what it means that something is *apparently* conceivable, but precisely what it means that something is conceivable *period*. Our position regarding these matters will become clear as we proceed.

5.4.4 Genuine and apparent conceivability: an intermediate position

In the two sections above, we have outlined two different attitudes towards the relation between conceivability (period) and the distinction between genuine and apparent conceivability. The view outlined in sections 5.4.1–5.4.2 identifies conceivability with genuine conceivability, and the view outlined in section 5.4.3 identifies conceivability with apparent conceivability. In this section, I shall outline a view according to which apparent conceivability, as we have outlined it, may *become* genuine conceivability. Such a view can, in large, be attributed to Tye (1995) and Yablo (1993). In my brief exposition, I shall focus on Tye (1995).[19]

In chapter 4, we considered Tye's views on conceivability in connection with the question whether there is any limit to Arnauldian skepticism. We noted that Tye thinks that it is absurd to doubt the following principle:

(39) If *x* believes or (honestly) claims that it appears to *x* that so-and-so, then it appears to *x* that so-and-so.

With respect to the notion of apparent conceivability, we saw no reason why we should not be able to extend this principle to cover appearances of conceivability. Thus:

(40) If *x* believes or (honestly) claims that it appears to *x* that *x* can conceive that S, then it appears to *x* that *x* can conceive that S—that is, *x* has *an apparent conception* that S, and S is *apparently conceivable* for *x*.

Tye argues that one is defeasibly warranted to believe that S is possible because of its apparent imaginability. With the introduction of appropriate new evidence, the warrant of this belief may be lost. (Here, Tye speaks of imaginability, not of conceivability, but I shall take this to be irrelevant to the present discussion. We further note that the same thesis appears in Yablo 1993: 34ff, and that it is similar to Kneale's (15) above.) In particular, Tye recognizes that the warrant may be lost due to the presence of possible defeaters based on the misdescription model. For example, my alleged conception of water being distinct from H_2O, which defeasibly warrants my belief that the same state of affairs is possible, can be defeated by the standard misdescription interpretation.

---

[19] Yablo's views on the matter have to some extent been discussed above. In section 1.3.1, we noted that Yablo believes that dialectical processes may resolve situations where conceivability intuitions are unclear or opposed, and in section 4.2.2, we discussed his models of modal error.

Here, Tye points to the possibility that there may be such "misdescription" defeaters available in *any* case of apparent conceivability.

Thus far, Tye's discussion has developed along familiar lines. At this point, however, where the usual reaction is often further despair, Tye is optimistic. He argues that if modal beliefs based on apparent conceivability cannot be defeated by a dialectical attempt to pin down the actual content of (the apparent) conception, there is reason to believe that apparent conceivability, in a given case, implies genuine conceivability. Tye considers a case where he imagines that there is a raccoon in the sitting room. Given (40), that state of affairs is *apparently conceivable* (for Tye). How do we get from apparent to genuine conceivability? Tye writes:

> Suppose that you ask me some further questions about what I have imagined, for example, whether I would still have counted the creature I imagined as a raccoon if its appearance had been changed dramatically so that it no longer looked like a raccoon at all. We might suppose here that I am asked whether painting a line down its back, attaching a bushy tail, and sewing inside it a sac of smelly liquid that it can squeeze and thereby produce a malodorous spray would turn the creature I am imagining into a skunk. If I answer negatively, then that is reason to suppose that I did genuinely imagine that there was a raccoon in the sitting room. If I answer positively, all I imagined was that there was a creature with a certain appearance (the one shared by raccoons) in the sitting room. In the former case, my act of imagination would continue to warrant the belief that a raccoon in the sitting room is metaphysically possible. (1995: 188)

Tye's concluding suggestion is that evidence from imaginability can be used to support claims about metaphysical possibility, but that "considerable care is needed in deciding just what has really been imagined" (1995: 188).

With the last remark, we are back on familiar tracks. As we have seen in the previous chapters, the problems connected to the task of "deciding just what has really been imagined" abound. The problem I think is particular to Tye's view is that he does not explain in virtue of *what* a raccoon, as opposed to a skunk, is imagined or conceived. In the absence of such an account, it is also difficult to see what details the dialectical process should focus on. However, the absence of such a *positive* account does not prevent a meaningful dialogue regarding whether someone has *failed* to imagine or conceive a particular object or situation. We shall provide an example of such a dialogue below, in addressing Tidman and Kind's claim to have imagined an Escher staircase.

5.4.5 Genuine and apparent conceivability: a dilemma

In general, I shall take it that the positions outlined in sections 5.4.1–5.4.3 present us with a dilemma which forces us to take a stand on what we shall take it to mean that something is conceivable, and how we shall interpret the conceivability thesis. We have already seen that the two positions identify conceivability (period) with different notions: with genuine and apparent conceivability, respectively. The *dilemma* consists in the fact that both identifications entail what appear to be intolerable consequences.

What is the intolerable consequence of the identifications made by the first position above? In order to explain this, we shall return to the exchange between Descartes and Arnauld. We note that neither Arnauld nor Descartes seems to think that Arnauld's demand for adequate knowledge presents the *main* problem for the conceivability thesis. "A created intellect," says Descartes, "may in fact possess adequate knowledge of many things." I believe that Arnauld would agree with Descartes that created intellects can possess adequate knowledge of many things, and that such knowledge can "easily" occur.

Now the intolerable outcome. Given that we sometimes—perhaps even often—possess adequate knowledge, *when* is our knowledge adequate? And how are we to know when it is? In order for a conceiver to know *beforehand*, regarding a particular application of the conceivability thesis, that *this* particular application is one of the valid applications, the conceiver needs not only to know all the essential properties of the thing in question, but also *that* he or she knows all the essential properties of the thing to which the conceivability thesis is applied. We have already indicated that Descartes thinks that such knowledge is unattainable (section 4.2.2). He writes:

> A created intellect […] though perhaps it may in fact possess adequate knowledge of many things, can never know it has such knowledge unless God grants it a special revelation of the fact. (CSM 2, Rep 4: 155)

Yablo (1993: 16) also argues that we cannot *know* that what we appear to conceive of is what we actually conceive of unless "the ideas employed are certifiable in advance as adequate—as embracing every essential property of their objects." In section 4.4.1, we saw that Kripke (1980) seems to accept this piece of bad news in general, but that he believes that there are special cases when we know that we know all the essential properties of a thing. In section 4.4.1, we focused on the mind-body thesis in particular, but now we are concerned with the validity of the conceivability thesis in general. And then, Yablo's bad news cannot be circumvented, at least not à la Kripke. This is at least Kripke's own contention.

In general, then, the requirement for genuine conceptions has intolerable epistemological consequences, in that it is not autonomously assessable, not open to introspection, whether you have managed to conceive what you sought to conceive. Can you ever be sure that you have (genuinely) conceived of what you apparently have conceived? Throughout this essay, we have mostly encountered negative answers to this question. On the other hand, we have now set things up so that there are ways to proceed even if this horn of the dilemma is chosen.

We accept the second horn of the dilemma by reasoning in the same way as Putnam does above. There may be various reasons for wanting to specify what it means that something is conceivable in terms of apparent conceivability, as Putnam does. For example, one may want one's concept of conceivability to track what we ordinarily mean by "conceivable." Another (more compelling) reason for wanting to specify this in terms of apparent conceivability is the epistemological benefits. In (40) above, we argued that if someone believes or (honestly) claims that he or she can conceive that so-and-so, then he or she *has an apparent conception* that so-and-so. Therefore, the question whether someone has apparently conceived of something or not can always be autonomously assessed by means of introspection, or by a simple question.

Here, the undesired outcome is on the possibility side. If we specify what it means that something is conceivable in terms of apparent conceivability, it will, at least to many conceivers, be conceivable that iron floats on water, that an actual proof P of theorem T is mistaken, that Phosphorus is not Hesperus, and that water is not H_2O. This is why Putnam (1975) cries out: "conceivability is no proof of possibility!"

Thus, our dilemma is this. We have decided to interpret the conceivability thesis in terms of strict possibility, because this is the relevant interpretation when it comes to philosophical argumentation (see chapters 3 and 4). Now, if we want to be certain that conceivability implies the designated form of possibility, we must put constraints on what it takes for someone to conceive of something that are overly rigorous. On the other hand, if we tend to the epistemological problem and loosen the constraints, identifying conceivability with *apparent* conceivability, our discussion above suggests that we can no longer be certain that conceivability implies the designated type of possibility. However, as we previously indicated, identifying conceivability with *apparent* conceivability does not preclude a modal dialogue such as that suggested by Yablo (1993) and Tye (1995), pertaining to the question whether someone has *failed* to imagine a particular object. (See the discussion pertaining to the

question whether it is possible to imagine an Escher staircase in the next section.)

5.5 What should we take it to mean that something is conceivable?

In this section, I shall not try to explain what conceivability really is, or what we must take conceivability to be, but rather what we should take conceivability to be, given that we think some desiderata on a philosophical notion of conceivability are more important than others.

Up to now, we have not mentioned many *concrete* counterexamples to the conceivability thesis. (We have rather been occupied with general considerations that speak for or against it.) This partly depends on the fact that we have not decided what our substantial notion of conceivability should be. During our discussion of the various distinctions presented in the previous sections, we have mentioned many substantial notions of conceivability. I enumerate them here for convenient reference. (In doing this, where there are a number of definitions that are obvious variants of each other, I shall omit all but the one that appears best formulated. In addition, in a case where a definition is an elaboration of, or restriction on another definition, I shall omit the latter, given that it has no virtues not retained by the corresponding elaboration.)

(3) To conceive of a proposition is to understand its meaning distinctly.

(4) To conceive of a proposition is to give some degree of assent to it, however small.

(5) A statement S is conceivable if it is consistent with the totality of conceptual truths, that is, if ¬S is not a conceptual truth.

(7) A statement S is ideally conceivable iff there is a possible conceiver *x* such that S is prima facie conceivable for *x* with justification that is undefeatable by better reasoning.

(8) A statement S is prima facie conceivable for a conceiver *x* when S is conceivable for *x* on first appearances. That is, after some consideration, *x* finds that S passes the tests that are critical for conceivability.

(9) A statement S is conceivable if S does not involve any apparent or explicable contradiction.

(11) A statement S is conceivable for a conceiver *x* iff *x* can imagine a possible world which he takes to verify S.

(14) A state of affairs is conceivable for a conceiver *x* iff *x* can form a mental image of the state of affairs in question.

(15) A proposition *p* is conceivable for me iff I have in mind no information which formally excludes *p*. (If I discover that ¬*p*, I withdraw my statement, because I then have information which entitles me to say that ¬*p*.)

(16) A statement S is conceivable for a conceiver *x* iff *x* can imagine acquiring evidence that would justify him/her in believing S.

(17) A proposition *p* is *strongly conceivable* for a conceiver *x* iff *x* intuitively sees that *p* is possible.

(18) A proposition *p* is *weakly conceivable* for a conceiver *x* iff *x* does not intuitively see that *p* is impossible.

(19) S is *positively conceivable* for a conceiver *x* when *x* can imagine a situation that verifies S.

(33) A statement S is prima facie positively conceivable for a conceiver *x* iff *x* can imagine a situation and judges, after some consideration, that if that situation had obtained, S would have been the case.

(34) A statement is ideally, positively, and secondarily conceivable for a conceiver *x* iff (i) *x* can imagine a situation that *x* takes to verify S, and (ii) there is no possible conceiver that can defeat *x*'s statement to this effect by means of better reasoning and/or more extensive empirical knowledge.

In the following, I shall discuss possible counterexamples to the conceivability thesis when interpreted in terms of some of these notions of conceivability. My aim is not to criticize some individual notion of conceivability, but to determine the general characteristics of notions of conceivability that are viable with respect to the desiderata we are about to present.

We begin with (15) above, which appears to be Kneale's interpretation of what it means that something is conceivable. The problem with (15) is obvious. First, whether something is conceivable or not will, according to (15), vary from conceiver to conceiver (given that the proposition considered is not an obvious a priori truth or falsehood), since different conceivers possess different information. Consider a conceiver with nearly no information at all. Since most propositions will be compatible with what he or she knows, most propositions will be conceivable for him or her, including impossible propositions such as the

proposition expressed by (22) above. Hence, there seem to be counterexamples to the conceivability thesis when we interpret it in terms of (15).

Another interpretation of what it means that something is conceivable, which is sometimes attributed to Hume (1978),[20] is (14) above. According to (14), there is a close relationship between conceivability and imagery. Now, as we briefly indicated in section 5.4.1, Tidman (1994) and Kind (1997) have argued that they can form a mental image of an Escher Staircase. By means of other sources of knowledge, we know that such a staircase could not possibly exist. Hence, Tidman and Kind argue, we have a counterexample to the conceivability thesis on the interpretation of what it means that something is conceivable in (14).

Pausing for a moment, we can see that these counterexamples are given relative to some particular substantial interpretation of what it means that something is conceivable. Hence, a counterexample to the conceivability thesis relative to some substantial interpretation of what it means that something is conceivable may not be a counterexample to the conceivability thesis relative to some other substantial interpretation. A trivial desideratum thus presents itself:

(D1) There should be no obvious counterexamples to the conceivability thesis on the favored substantial interpretation of conceivability.

This desideratum should not be interpreted as saying that there should be no counterexamples to the conceivability thesis *period*, but rather as saying that, before we consider elaborate notions of conceivability (in terms of the distinctions set forth in the previous sections), we should attempt to find a basic notion of conceivability according to which conceivability judgments are not susceptible to obvious modal error.

When considering proposed counterexamples to the conceivability thesis, we must also evaluate whether it is a genuine counterexample, or if it is possible to cast doubt on it. Some counterexamples seem unquestionable, such as the considerations we presented against Kneale's definition. However, we may sometimes vindicate a substantial notion of conceivability by means of calling the counterexample into question. For example, as I have indicated in section 5.4.1, I believe that it is possible to cast doubt on the alleged counterexample to (14) presented by Tidman and Kind. The object we know to be impossible is the object depicted in Escher's picture: the staircase. It can plausibly be argued that the object which Tidman and Kind *actually* have imagined, when they claim to have imagined the object in the Escher picture, is not that object, but the picture

[20] See Van Cleve (1983) and Tidman (1994).

itself. Assume that Tidman and Kind actually have misidentified the object of their imaginings in this way. If this is the case, the counterexample to the conceivability thesis disappears: the imagined thing is surely possible, since it actually exists.

How could you support the claim that Tidman and Kind have misidentified the object of their imaginings in the way described? Consider the properties we normally associate with imaginings of various objects. For example, if I imagine a three-dimensional object, such as my bedroom, it seems as if I am free to assume any perspective on the object imagined. I can, for example, imagine standing at the foot of the bed, or at the window, and obtain the corresponding perspective on my bedroom. When I imagine my house, I can imagine it from in front, from aside, from an angle, and so on. This is how we imagine three-dimensional objects. Now, Tidman and Kind claim that they can imagine the *object* depicted in the Escher picture. If so, they should be able to imagine it from any perspective they want. But can you really do this with the object depicted in the Escher picture? The answer seems to be no. When imagining the object depicted in the Escher picture, you are stuck with the perspective of the *picture*. But then, if your imagining shares the perspective of the picture, is it not reasonable to conclude that you are in fact imagining the *picture itself*, and not the object in the picture?

We shall leave it open what the object of Tidman and Kind's imaginings actually is. We cannot raise doubt as to whether they have *apparently* imagined an impossible staircase. However, the above line of reasoning shows that it is possible to raise doubt regarding whether they have *genuinely* or *ideally* imagined such a staircase. In cases where the conceiver is presented with such arguments, he or she might retract the claim to have imagined the designated object.

As we proceed to consider alternatives to (14) and (15), we will find that there are many substantial notions of conceivability according to which the conceivability thesis withstands the counterexamples we have considered so far. Consider for example Balog's (5) above. (5) is not affected by the counterexample to (15), since there is no mention of conceivers and their epistemic states. It also seems as if the Escher counterexample to the conceivability thesis can be avoided if we replace (14) with (5), since there are all reasons to suspect that any description of the Escher staircase would be conceptually *inconsistent*.

However, even if there are no obvious counterexamples of the kind we have considered thus far to the conceivability thesis as understood in terms of (5),

there are other considerations that speak against it. First of all, as we have pointed out earlier, conceptual consistency does not imply genuine possibility. But even if we modified (5) to amend this (say, by adding a clause that S must not contradict any truth about essences), there is still something deeply problematic with this type of definition. In section 5.3 above, we considered absolute and relative (and relational) notions of conceivability. On absolute notions of conceivability such as (5), conceivability is a property of statements. I believe that virtually the same considerations that speak against the attempt to identify conceivability with genuine conceivability also speak against the "property" interpretation. Suppose that we understand conceivability to be a property of statements. Now we must find a method for determining whether a statement has this property (the property of being conceivable). We become involved in a kind of regress: the conceivability or non-conceivability of a statement was supposed to inform us about the modal properties of this very statement. Now we have to devise a method for finding out whether a statement has the property of being conceivable. This does not solve the epistemological problems: the question *how we can know* that something is conceivable reemerges, as it does in the attempt to identify conceivability with genuine conceivability.

I want to conclude this chapter by taking a stand in some central issues. First, I suggest that conceivability should not be viewed as a property of statements, but as a *relation* that may or may not obtain between a conceiver and a statement. (We will consider the third alternative, according to which conceivability is a *relational* property of statements below.) My suggestion is based on epistemological considerations. If we decide to understand conceivability as a property of statements, the conceivability thesis will be deprived of its epistemological relevance. My idea is that given that conceivability is described as a *relation*, a definition of what it means that something is conceivable must *at least* specify what it means that the conceiver stands in this relation to that which is conceived (see section 5.2.3). It seems to me that such a specification must be given at least partly in terms of intrinsic properties of the conceiver, such as mental states.

These considerations suggest the following desideratum:

(D2) A definition of what it means that a statement S is conceivable for a conceiver *x* should specify, in mental or cognitive terms, what it means that a conceiver *x* stands in the relation "S is conceivable for *x*" to S.

I shall address the question of how we should respond to *relational* notions of conceivability in relation with Chalmers' notion of *ideal* conceivability in (7), and the more complex definition of conceivability (35), which is based on the notion of ideal conceivability. Above, we noted that Chalmers believes that there are two versions of the conceivability thesis that are valid, namely (31) and (32). Here, we shall only consider (32), since this is the version of the conceivability thesis that is relevant for our purposes. This version of the conceivability thesis states that *ideal secondary conceivability entails secondary possibility*.

It is important that the conceivability in question is of the ideal variety for the conceivability thesis (32) to be plausible. Chalmers writes: "In any case, if we are looking for a notion of conceivability such that conceivability tracks possibility perfectly, we must focus on ideal conceivability" (2002a: 160). I shall explain why I believe that ideal notions of conceivability may be unacceptable as interpretations of what it means that something is conceivable *simpliciter*. Chalmers does not attempt to explicate what good or undefeatable reasoning consists in, or what counts as a cognitive limitation to be "idealized away from," since he believes that any such attempt will turn out to be "open-ended and incomplete" (2002a: 148). However, if we assume (as Chalmers seems to do) that it is safe to "idealize away" from certain cognitive limitations, we will perhaps end up with definitions of what it means that something is conceivable which are useless to human, limited conceivers in philosophical practice. No matter how accidental we believe a general limitation of the human intellect to be, it is still *a limitation of the human intellect*. A conceiver who does not share our cognitive limitations can perhaps not be described as one of us in the relevant sense, and even if something is conceivable to such a conceiver, how are we to know? In other words, a version of Arnauld's skeptical question can be raised if someone claims that something is ideally conceivable: how can you (ever) know that your conception is ideal?

In one sense, we keep returning to the same problem. We began with the distinction between genuine and apparent conceivability, which is grounded in modal facts about the things that we conceive. Later, we presented the distinction between ideal and prima facie conceivability, which is grounded in requirements on the cognitive processes on which conceivability judgments are based. However, the dilemma seems to remain: genuine or ideal conceivability seems to entail possibility, but apparent or prima facie conceivability appears not to do so. On the other hand, if conceivability is identified with apparent conceivability, we can autonomously assess whether something is conceivable

or not, whereas we cannot thus assess whether something is genuinely or ideally conceivable. In order to bridge these gaps, we would need to be able to infer genuine conceivability, and thus possibility, directly from apparent conceivability. However, we keep ending up with this fact:

(41) It is not the case that apparent conceivability implies conceivability *or* it is not the case that conceivability implies possibility.

That is, either conceivability is identified with apparent conceivability, in which case conceivability does not imply possibility, or conceivability is identified with genuine conceivability, in which case conceivability implies possibility, but in which case conceivability facts cannot be autonomously assessed. Either alternative is unsatisfactory. Much of the remainder of this essay will concern the question if, and to what extent, any of these gaps can be bridged.

As we noted earlier, Arnauld never denies the conceivability thesis. In other words, he never denies the

(42) conceivability implies possibility

part of (41). What he denies is the

(43) apparent conceivability implies conceivability

part of (41). By contrast, Putnam accepts (43), but rejects (42). Chalmers' distinctions are epistemological, and are therefore presented in different terms. Nevertheless, he presents us with roughly the same situation as Arnauld. Chalmers argues that given the distinctions between ideal and prima facie conceivability, positive and negative conceivability, and so on, it is "relatively easy to classify potential gaps between conceivability and possibility" (2002a: 159). He begins with:

(44) Prima facie conceivability is an imperfect guide to possibility.

On the other hand, Chalmers believes that:

(45) (A form of) ideal conceivability entails possibility.

However, he notes that:

> Prima facie secondary conceivability judgments can go wrong as a guide to secondary [metaphysical, genuine] possibility when a subject is misinformed about the relevant nonmodal facts, and perhaps when an incautious subject is merely ignorant of these facts. (2002a: 173)

In other words:

(46) It is not the case that prima facie conceivability implies ideal conceivability.

How should we respond to these claims? Should we maintain the validity of the conceivability thesis, or should we assert the epistemic transparency of conceivability data, and provide an account of conceivability, according to which conceivability can be autonomously assessed by the conceiver? Of course, to an *ideal* conceiver, who for example possesses adequate knowledge of all things or sufficient non-modal knowledge about everything, even ideal conceivability data are epistemically transparent. But what should *we* do, and how should *we* use (and respond to the use of) the conceivability thesis in philosophical argumentation?

I suggest that we should opt for epistemic transparency of conceivability data, that is, for the truth of the transparency thesis (1). This suggestion amounts to the following final desideratum. The desideratum presented in (D2) is that a definition should describe what it means that a conceiver *x* stands in the relation "*y* is conceivable for *x*" to *y*. To this desideratum, we add the following:

(D3) The conceiver *x* should (in the relevant cognitive sense) be like one of us.

I believe that this conception of the problem is similar to that of Descartes and Arnauld. First, the discussion between Descartes and Arnauld concerns a strict (or metaphysical) notion of possibility. Something is possible if God can bring it about, so clearly, modal facts obtain independently of human conceivers and their cognitive abilities. Secondly, *conceivability* is clearly an epistemological notion. Whether something is conceivable or not depends on the psychological endowment of the conceiver. Suppose that a certain statement is in fact possible. However, this statement is so complex that no ordinary imperfect conceiver can conceive of it (that is, conceive of it as true in some imagined situation). Perhaps only God or some other ideal conceiver can conceive of it. Then, although this statement is conceivable for God and for ideal conceivers, and therefore conceivable in one sense, it is not conceivable *for us*, and therefore not conceivable in the sense that Descartes and Arnauld—and we—are interested in.

The modal epistemologist must make a choice. I have proposed that in the choice between a universally valid conceivability thesis and epistemic transparency of conceivability data,[21] we should choose the latter. If we dissociate conceivability and *apparent* conceivability, we have just replaced the question "How can we know whether something is possible?"—which we

[21] The term "epistemic transparency of conceivability data" is adopted from Almog 2002a: 18.

originally wanted to answer with reference to conceivability facts—by the equally difficult question "How can we know whether something is (genuinely) conceivable?"

However, to identify conceivability (period) with *apparent* conceivability, is not to say that apparent conceivability judgments cannot be contested, for example in the way we have contested Tidman's and Kind's apparent conceptions of an Escher staircase above.

CHAPTER 6

MODALITY

6.1 Introduction

In this chapter, I will be concerned with possible ways of understanding the notion of *strict* possibility introduced in chapter 3. In that chapter, I distinguished between *objective* possibility, such as logical and metaphysical possibility, and *subjective* possibility, such as epistemic and doxastic possibility. Relative to this distinction, I introduced the concept of strict possibility as follows: a possibility notion N is *strict* iff the following holds with respect to it:

(1) A statement S is possible with respect to N iff the world could have been such that S was true.

In this thesis, I have assumed that truths about how the world could have been are objective.[1] As I have previously pointed out, the reasons for using the term 'strict possibility' instead of the more common Kripkean 'metaphysical possibility' are twofold. First, I want to dissociate the notion of strict possibility from well-known Kripke-inspired views about what *makes* certain statements metaphysically possible (necessary). Secondly, in the present chapter, I intend to use the term 'metaphysical possibility' in a more detailed sense than that of "truth in at least one possible world."

The question I shall raise in the present chapter is this: what types of facts are (ultimately) responsible for the truth of true modal statements? When it comes to strict possibility, this question amounts to the following: what types of facts determine the possible ways in which the world could have been? Many philosophers agree that there are objectively true modal statements, and their notions of possibility may even be extensionally equivalent. However, they need not agree on what *kinds* of fact determine the possible ways in which the world could have been. For example, as will to some extent be made clear below, some philosophers think that that it is primarily conceptual facts that

---

[1] For any strict notion of possibility, there is a corresponding notion of *strict necessity*: S is strictly necessary iff the world could not have been such that ¬S were true.

determine the possible ways in which the world could have been, while others think that this is determined by properties of the things themselves.

Traditionally, philosophers have focused on logical notions of modality, and distinguished between *narrow* and *broad* logical possibility and necessity (see section 6.2 below). It is traditionally assumed that all logical modal truths are a priori. However, since Kripke (1980), philosophers have also begun to distinguish between statements that are modal truths in *the metaphysical sense* and statements that are logical modal truths. The basic idea behind this distinction is that metaphysical modal truths are truths that obtain in virtue of facts about the identity of things (broadly conceived),[2] whereas broadly logical modal truths are truths that obtain in virtue of relations between concepts. An important feature of Kripke's theory is that there are metaphysical modal truths that can only be known a posteriori, such as "□(Water = H_2O)" and "□(Hesperus = Phosphorus)." These modal statements are a posteriori since they require the empirical knowledge that Water is H_2O, and that Hesperus is Phosphorus.

In this chapter, I shall outline an account of strict modality according to which all modal truths have their basis in conceptual truths. More specifically, I shall argue that metaphysical modal truths are established by applying a priori principles of broad logical modality to nonmodal a posteriori truths. The idea that these concepts are interrelated in this way is not original,[3] but it is seldom spelled out in detail. I believe that a clarification of the relations between these concepts provides the best possible starting point for a conceptual interpretation of metaphysical modality.

I begin by giving an account of broad logical modality in section 6.2, and of metaphysical modality in section 6.3. I shall sometimes speak of both necessary truth as well as modal truth in general (comprising necessary truth and possible truth), but I shall focus on *possible truth*, in order to facilitate the subsequent discussion of the relation between conceivability and possibility in chapter 7. In

---

[2] Cf. Fine 2002: 254; 1994: 8–9. The distinction made here does not, however, correspond to Fine's distinction between "the new school" and "the old school." According to Fine, philosophers of the "new school" maintain that there is only one type of necessity, that which obtains in virtue of the identity of things (broadly conceived). Fine calls this type of necessity *broad logical* or *metaphysical* necessity. By contrast, philosophers of the "old school" maintain that the only type of necessity there is is that of *narrow logical* (or *formal*) necessity (for this notion, see below). Thus, where we distinguish between 'broadly logical' and 'metaphysical' necessity, Fine uses these terms to refer to the same type of necessity.

[3] For similar ideas, see Kripke (1980: 3); Baldwin (2002: 18–20); Jackson (1998: 59); van Inwagen (1998: 82), and Weigel (2000: 218).

section 6.4, I turn to a critical discussion of the notion of metaphysical modality. This discussion is concluded in section 6.5, where I deal with different views on metaphysical possibility that are compatible with our conclusions in section 6.4. In section 6.6, I summarize the discussion of this chapter, and make some remarks that will function as a point of transition to our concluding discussion in chapter 7.

6.2 Broad logical modality

There is more than one form of logical modality. On the one hand, there is the notion of *narrow* logical modality. We shall say that a statement S in a language $\mathcal{L}$ is *narrowly logically possible* iff there is an interpretation (model) $\mathcal{M}$ (for $\mathcal{L}$) in which S is true. A statement S is *narrowly logically necessary* iff S is true in every interpretation. Intuitively, a statement S is narrowly logically necessary if S is true strictly in virtue of the *logical* concepts involved in S. On the other hand, there is the notion of *broad* logical modality. To provide a satisfactory definition of the concept of broad logical modality is problematic. When it comes to the question what such a definition should look like, there are two approaches in the literature. First, it is possible to provide a definition of broad logical possibility in terms of conceptual truth (see Chalmers 1996: 52), but it is also possible to provide an epistemological definition in terms of a priori knowability (see Bealer 2002: 71 and Boghossian 1996). There are problems with both accounts. For example, Quine's views in his "Two Dogmas of Empiricism" (1961 [1951]) are in conflict with both approaches. In particular, his criticism of the analytic/synthetic distinction is in conflict with the claim that some statements are necessary in virtue of the meaning of the component terms. Likewise, Kripke's (1980) claim that there are a priori contingent truths is perhaps in conflict with the attempt to define broad logical necessity in terms of apriority. For example, suppose that M is the standard meter in Paris, and that the reference of the term 'one meter' is fixed by "the length of M at a certain time *t*." Kripke (1980: 54–7) notes that the statement "M is one meter long at *t*" is a priori, since M is used to fix the reference of the term 'one meter'. However, Kripke notes, the statement "M is one meter long at *t*" is not a necessary truth, since M could have been longer, or shorter, at *t*. If Kripke is correct, there are statements, such as the above, that are a priori true but not metaphysically necessary.[4]

[4] On the other hand, the statement "One meter = the *actual* length of M at *t*" seems to be both a priori true and metaphysically necessary.

Despite the various difficulties with such a definition, we shall define broad logical possibility in terms of conceptual truth. As Chalmers (1996: 52) notes, the main problems with the notion of conceptual truth are these:

(i) Most concepts do not have definitions giving necessary and sufficient conditions.

(ii) Most apparent conceptual truths are in fact revisable, and could be withdrawn in the face of sufficient empirical evidence.

(iii) Considerations about a posteriori necessity show that conditions of application across possible worlds cannot be known a priori.

Our discussion in sections 6.4–6.6 will relate to (iii). With respect to (i), one could argue as follows (given the theory we are about to outline). From a conceptual realist perspective, even if (i) is true, this need not be a problem for the notion of conceptual truth. Conceptual truths obtain, on a realist view, in virtue of relations between concepts. It may be that our *definitions* of certain concepts do not provide conditions that enable us to determine whether a certain statement is a conceptual truth or not, but this is a problem that pertains to our definitions, not to the concepts themselves. In other words, since at least some concepts are assumed to be independent of the definitions that are supposed to capture them, the notion of conceptual truth as a relation between concepts is unaffected by (i).

With respect to (ii), one could argue as follows. According to a moderate version of conceptual realism, not all concepts need to be immune to empirical evidence. Such evidence may suggest that people seem to be able to conceive the impossible. If we want to preserve the link between conceivability and possibility, we may, given such evidence, revise our concept of conceivability (or our concept of possibility). However, there may be concepts that are immune to any empirical evidence. One such concept is perhaps the concept of identity (and concepts that are closely related to the concept of identity). In other words, conceptual truths pertaining to the concept of identity, such as (5) and (6) below, are, according to this view, not revisable. Instead, we think of them as constitutive of *any* possible conceptual schema.

We shall say that a statement S is broadly logically possible iff ¬S is not a conceptual truth. I will not attempt to explain what a concept really is. Nevertheless, I will make some assumptions regarding concepts. Firstly, I shall assume that the *meanings* of terms in our language can be identified with concepts. Hence, I will speak of terms (‘chair’) and the concepts associated with these terms (*chair*). Secondly, I shall take it to be possible to conduct what

is commonly known as "conceptual analysis."[5] In one sense of "conceptual analysis," it consists in the attempt to explicate the concept associated with a certain term. In this sense, we could say that a statement S is broadly logically possible iff ¬S cannot, even in principle, be identified as a conceptual truth by means of conceptual analysis.

Before we proceed, we need to discuss Boghossian's (1996) and Harman's (1967) criticism of the concept of analyticity we have employed here. On our understanding, a statement S is an analytical (or conceptual) truth if S is true in virtue of its meaning. Boghossian refers to this concept of analyticity as the *metaphysical* concept of analyticity. He contrasts it with the *epistemological* concept of analyticity, according to which a statement S is an analytical truth if the mere grasp of S's meaning is sufficient for being justified in holding S true.

Boghossian points out that for each statement S, analytical or not, in order for S to be true, *it must be the case that* S. In other words, it holds for each true statement S—even for analytically true statements—that the truth of S also depends on the *fact* that S is the case. In this context, Boghossian quotes Harman (1967), who asks the following question:

> [W]hat is to prevent us from saying that the truth expressed by "Copper is copper" depends in part on a general feature of the way the world is, namely that everything is self-identical? (Harman 1967: 128)

In so far as I have understood Boghossian and Harman correctly, I think that their criticism is highly questionable. Harman seems to argue that the truth of the statement

(2) Copper = copper

is in part due to a certain *fact*, namely the fact that the world is such that everything is self-identical, that is:

(3) $\forall x(x = x)$.

On the usual understanding however, (3) is not just any old fact about *the world*, but a (narrow) *logical truth*, and as such, an analytical truth and a "general feature" of *each* possible world (which is to say that $\Box\forall x(x = x)$ holds). Thus, on my view, it is misleading to describe (3) as a general feature of *the world*, as Harman does, when it is in fact a general feature of *each possible world*. (A "general feature" of *the world* is, perhaps, the law of gravity.) In conclusion,

---

[5] Different activities have been associated with "conceptual analysis" through the history of philosophy. One could respond to this diversity in a pluralistic fashion, and suggest that "conceptual analysis" should be understood as comprising all of these activities. We could then say that there are *many aspects* of conceptual analysis.

Harman and Boghossian are correct in that the truth of (2) is dependent on (or follows from) the truth of another statement, namely (3), or, if you will, the *fact* that (3) is true. However, (3) is itself an analytical truth, and there is no reason why a statement S_1 could not, for its truth, depend on the truth of another analytical truth S_2, and still be regarded as an *analytical* truth. (After all, the statement "(3) implies (2)," that is, where '*a*' is a name of copper,

(4) $\forall x(x = x) \rightarrow (a = a)$

is itself an analytical truth.)

6.3 Metaphysical modality

Roughly following Fine (2002), I shall take modal metaphysical truths to be truths that obtain in virtue of the identity of things, broadly conceived. What does it mean that a statement is necessary in virtue of the identity of things "broadly conceived"? One way to interpret Fine is the following. First, we take some statements to be necessary in the intended sense in virtue of the identity of things *narrowly conceived*. Kripke (1980: 3) points out that from the thesis that all things are necessarily self-identical, that is:

(5) $\forall x \, \Box(x = x)$

and *Leibniz' law*:

(6) $\forall x \forall y((x = y \wedge Fx) \rightarrow Fy))$

it follows that all identical things are necessarily identical:

(7) $\forall x \forall y(x = y \rightarrow \Box(x = y))$.

From (7), using the rule of universal specification, we obtain:

(8) (Phosphorus = Hesperus) $\rightarrow \Box$(Phosphorus = Hesperus).

Here, we have assumed that *x* and *y* in (7) can be substituted by proper names such as 'Phosphorus' and 'Hesperus' via the rule of universal specification. Now, we know from empirical investigations that the following identity statement is true:

(9) Phosphorus = Hesperus.

By modus ponens, it follows from (8) and (9) that:

(10) $\Box$(Phosphorus = Hesperus).

Secondly, some statements can be described as necessary in virtue of the metaphysics of essence, or, in virtue of the identity of things *broadly conceived*.

From the notions of essence we shall consider, it is clear that the metaphysics of essence is part of the metaphysics of identity. According to one definition of essence, a property *F* is essential to an object *a* if it is necessary that, if *a* exists, then it has the property *F*:

(11) $\Box(\exists x(x = a) \rightarrow Fa)$.

There are also non-modal definitions of what it means that a property is essential to an object. As Fine (1994: 2) notes, Locke understood the essence of a thing to be "the being of any thing, whereby it is what it is," and this clearly seems like a non-modal definition of essence. (Actually, Fine (1994) and Almog (1991) have argued against the attempt to explain essence in terms of modality, as is done in (11). More will be said about this in section 6.5.2.)

What would be an example of a statement that is necessarily true in virtue of the identity of things *broadly conceived*? Fine claims that

(12) Kit Fine is a person

is such an example (cf. 2002: 254). Notice that the necessity of (12) does not depend on the identity of things *narrowly conceived*, that is, (12) is not a statement of the form "$a = b$" (where '*a*' and '*b*' are singular terms). On the contrary, (12) is a statement of the type "*Fa*," where *F* is an essential property of *a*, that is, where *F* is a property such that *a would not have been a without it*. Thus, (12) is necessary—in virtue of the identity of things *broadly conceived*.

6.3.1 How to establish metaphysical modal truths

The thesis that metaphysical modal truths are established by applications of the model we shall outline below has been suggested, more or less directly, by many philosophers, including Kripke (1980: 3); Baldwin (2002: 18–20); Jackson (1998: 59); van Inwagen (1998: 82), and Weigel (2000: 218). This model was briefly suggested in section 2.5, in connection with our discussion of van Inwagen (1998).

(i) First, there is an (a priori) *broadly logical principle of modality*, which, in the typical case, states conditions for identity (broadly conceived). Such a principle might invoke any of the definitions of essence considered thus far, but it might also invoke the physical composition of things, such as "it is not possible for a thing or substance to have a physical composition that is different from the physical composition that it actually has" (cf. Weigel 2000: 218; van Inwagen 1998: 82; Jackson 1998: 70). Or it might appeal to origin or historical factors, such as "if *x* and *y* are the gametes of *z*, then *x* and *y* are necessarily the gametes of *z*" (cf. Baldwin 2002: 19).

(ii) Secondly, there is an observation of a certain (non-modal) fact concerning the identity of things, such as the fact that water is H_2O, that $2 + 2 = 4$, or that a person comes from certain gametes.

(iii) Thirdly, the relevant a priori principle of possibility is applied to the nonmodal fact, and a modal inference is made: "The substance water could not have had a physical composition that is different from the actual physical composition of water (H_2O)"—in short, $\Box$(Water = H_2O), or, "Since all identical things are necessarily identical, it is necessary that $2 + 2 = 4$." Schematically:

The mathematical case:

Broadly logical principle of modality: (7).	(Premise)
Mathematical truth: $2 + 2 = 4$.	(Premise)
Metaphysical modal truth: $\Box(2 + 2 = 4)$.	(Conclusion)

The "Water = H_2O" case:

Broadly logical principle of modality: (7)	(Premise)
A posteriori truth: water = H_2O	(Premise)
Metaphysical modal truth: $\Box$(Water = H_2O).	(Conclusion)

At this point, one might wonder: what is the difference between the allegedly "metaphysical" necessity of "$2 + 2 = 4$" in the mathematical case and the type of necessity that we described above as *broadly logical*? The answer is that there is no difference: "$2 + 2 = 4$" is indeed a broad logical necessity. But then, since both the mathematical case and the water/H_2O case are built on the same template, is "Water = H_2O" also a broad logical necessity? No, it is not, because it is not a conceptual truth.

There thus seems to be some kind of confusion involved in our attempt to classify the different types of modality: there seems to be a partial overlap between broad logical necessity and metaphysical necessity. However, the confusion is cleared away once we (trivially) recognize that the "template" on which the mathematical case and the water/H_2O case was built is applied to *two different types of non-modal truths*. In the mathematical case, the template is applied to an a priori truth ("$2 + 2 = 4$"), and in the water/H_2O case, it is applied to an a posteriori truth ("Water = H_2O"). The difference between the mathematical case and the water/H_2O is thus not that there are two kinds of modality involved. In both cases, '$\Box$' has the same reading. Rather, the difference is *epistemological*. In the non-mathematical case, one of the premises can only be known by means of empirical investigations.

All modal principles employed by the template are themselves broad logical necessities, and are as such a priori conceptual truths. When they are applied to

an a priori truth, the resulting modal conclusion is itself an example of broad logical modality. (The modal inference and its premises are a priori through and through.) On the other hand, when the template is applied to an a posteriori truth (such as "Water = H_2O," "Phosphorus = Hesperus," "Gold has atomic number 79," or "Cats are animals"), the resulting modal inference is always an example of what we call a "metaphysical" modal truth.

According to the view of the relation between broad logical modality and metaphysical modality that I have outlined here, metaphysical modal truths are established by applying broad logical truths to a posteriori truths about empirical facts. In other words, for there to be any characteristically *metaphysical* modal truths over and above broadly logical modal truths, such as the metaphysical modal truths mentioned above, there *must* be some physical facts to which the broadly logical principles can be applied.

6.3.2 Bootstrapping metaphysical modal truths

If it is true, as we argue, that metaphysical modal truths are dependent on a posteriori truths, it seems as if they could have been different given that other a posteriori truths would have obtained. Suppose, for the sake of argument, that certain physical facts would have been different, to the effect that some other possible world would have been the actual. We can easily imagine circumstances under which, for example, the stuff in lakes were not H_2O, but some other substance (XYZ)—or circumstances in which we found out that the creatures we refer to as "cats" turned out to be Martian spies or automata. In such circumstances, would there not also be different metaphysical modal truths? Let us apply our model to some empirical truths about these circumstances:

The "Water = XYZ" case:

Broadly logical principle of modality: (7).	(Premise)
A posteriori truth: Water = XYZ.	(Premise)
Metaphysical modal truth: □(Water = XYZ).	(Conclusion)

The "Cats are automata" case:

Broadly logical principle of modality: roughly, (11)	(Premise)
A posteriori truth: Cats are automata.	(Premise)
Metaphysical modal truth: □(Cats are automata).	(Conclusion)

According to Kripke, the above piece of reasoning is invalid. First, he would argue that a world in which the imagined scenarios obtained is not a world in

which the relevant a posteriori facts are different. In such a world, there may be watery stuff in the lakes that is not H_2O, and cat-like creatures that are automata. Nevertheless, Kripke would argue, none of these facts about that world are such that the metaphysical modal truths about the *actual* world—"Water is necessarily H_2O" and "Cats are necessarily animals"—would fail to hold with respect to that world, simply because the watery stuff in the lakes in that world is not water, and the cat-like creatures in that world are not cats. For any world to contain water or cats, it has to contain the stuff and the creatures of the *actual* world, and in the actual world, the relevant stuff is H_2O and the relevant creatures are animals.

Note that the last claim—which is central to the notion of metaphysical modality—can be disambiguated in at least two ways. First, the claim can be interpreted in what I shall call the *indicative* (or *demonstrative*) way. On this interpretation, the claim should be understood as follows:

(13) Any stuff in any possible world is water iff it is the same stuff as *this* (some actual water is pointed out).

However, the claim can also be interpreted in the *qualitative* way. On this interpretation, what it is to be water is defined directly in terms of (essential) properties of the stuff pointed out as "*this*" in (13):

(14) Any stuff in any possible world is water iff it is H_2O.

I believe that Kripke's argument should be interpreted in terms of (13). In other words, some kind of (primitive) "sameness" relation is first assumed, and it is *afterwards* determined just what this relation consists in by empirical investigations into the nature of, in this case, water. (For the idea of a "sameness" relation, see also Putnam 1975. What such a relation could consist in will be further discussed in section 6.4.) Suppose that such investigations veridically tell us that the chemical composition of water is H_2O. Then, if we accept that "being H_2O" is a *sufficient condition*, and not merely a *necessary condition*, for being water or "waterhood"—we can *replace* the indicative condition stated in (13) with the qualitative condition stated in (14). Of course, on Kripke's view, (14) has always been true, independently of our pointing out some particular substance as "Water."

Now, this is how the metaphysical modal truths about the actual world are "bootstrapped": if it holds for each possible world that something is not water unless it is H_2O, and not a cat unless it is an animal (and so on), this entails that familiar a posteriori truths about the *actual* world such as "Water is H_2O," "cats are animals," and so on, obtains in each possible world *w* irrespective of the

empirical facts about *w*. In other words, even if the world had been different in the ways we outlined above, it would still have been true that water is necessarily H_2O and cats are necessarily animals.

6.4 Rigid designation and "sameness"

Above, we have outlined a model of how metaphysical modal truths are established, and we have seen that even if such truths are dependent on a posteriori truths about actual physical facts (which could have been different), they could not themselves have been different. We further noted that the argument against the intuition that metaphysical modal truths could have been different relies on some sort of "sameness" relation—a relation that allows Kripke to argue that XYZ water is not water and automaton cats are not cats. In sections 6.4 and 6.5, I shall conduct a detailed discussion of the question of what this "sameness" relation is. We shall begin this discussion by considering a fallacious application of the model we have outlined above. Consider the following:

The "The watery stuff = H_2O" case:

Broadly logical principle of modality: (7).	(Premise)
A posteriori truth: The watery stuff = H_2O.	(Premise)
Metaphysical modal truth: □(The watery stuff = H_2O).	(Conclusion)

The premises in this argument are true, but the conclusion is false: it is not necessary that watery stuff is H_2O (that is, it is not true in all possible worlds that watery stuff is H_2O), because the watery stuff in Twin World is XYZ (I discussed the Twin Earth example in sections 4.4.1 and 5.3.2 above). Then, to what class of a posteriori truths can the model be applied? Kripke's (1980) well-known answer is that the model can only applied to identity statements that exclusively involve *rigid designators*, that is, singular terms that pick out the same objects in each possible world. The second premise in the "Water = H_2O" case above is such an identity statement, since both 'Water' and 'H_2O' are rigid designators. However, the second premise in the above inference is not such an identity statement, because 'the watery stuff' is not a rigid designator (that is, 'the watery stuff' does not pick out the same substance in all possible worlds).

What does it mean that a rigid designator, such as 'water', picks out the *same* entity in each possible world? Conversely, what does it mean that a non-rigid designator, such as 'the watery stuff', picks out *different* entities in different possible worlds? At first glance, it seems as if the question we raise here pertains to *the problem of trans-world identification*: if we suppose that objects

exist in more than one possible world, in virtue of what are objects re-identified as the same objects?[6] In *Naming and Necessity*, Kripke rejects the problem of trans-world identification as a pseudo-problem (1980: 15, 19, 52–3). He argues that we do not begin with the possible worlds themselves, and then ask for criteria of trans-world identification. Rather, he says, "we begin with the objects, which we *have*, and can identify, in the actual world. We can then ask whether certain things might have been true about the objects" (1980: 53). I believe that Kripke is correct in his rejection of the problem of trans-world *identification*. In some sense, we have the objects, and we can speak about them in counterfactual situations without having to worry about whether we have lost sight of the object we were talking about in the first place (say, Nixon). However, the problem of trans-world *identity*, and our question(s), remain: in virtue of what is an object (Nixon) the same object throughout the set of possible worlds that it exists in? In virtue of what does 'Nixon' refer to the *same* object in each possible world? Observe that our question does not pertain exclusively to the problem of trans-world identity—it is more general than that: what does it mean for an entity *x* to be the *same* entity as an entity *y*? In other words, what does the "sameness" relation that we discussed earlier consist in?

Consider the passage from *Naming and Necessity* that we cited in section 5.4 (Kripke 1980: 46–7). What seems to be clear from this passage and the above (1980: 53) is that Kripke thinks that we (in some sense) "fix" the object of discourse (Nixon) *prior* to modal considerations ("We […] simply consider *Nixon* and ask what might have happened to *him* had various circumstances been different"[7]). Our discussion of the problem of trans-world identification above supports this interpretation, because the alternative interpretation is that the object of discourse is "fixed" *by means of* modal considerations, in which case the problem of trans-world identification would reappear. (The problem of trans-world identification would reappear because we would, on the alternative interpretation, start out with no intuitions about what it is for an object to be the *same* object in counterfactual cases—this is for the counterfactual cases themselves to decide.)

In other words, the question of sameness seems to be prior to the question of whether a term is rigid or not. (For its intelligibility, the notion of rigid designation presupposes that it makes sense to say of an object *x* in a world *w* that it is the same object as *y* in *w**.) According to our discussion so far, Kripke provides the following account: *first*, the object of discourse (Nixon) is fixed,

---

[6] For a detailed discussion of this problem see, for example, Chisholm 1979.

[7] Kripke 1980: 47.

and *then*, modal considerations are brought to bear on the question of whether some singular term ('Nixon', 'the president of the United States in 1970') rigidly denotes that object. Thus, we have made some progress, in that the discussion in the last few paragraphs has suggested that we can address our problem; in virtue of *what* is an object the same object throughout the set of possible worlds that it exists in?—just by considering how we pick out the objects of discourse in the *actual* world.[8] How is this done? In particular, how is this done according to Kripke?

6.4.1 Kripke on objects and sameness

The problem is that when we turn to what Kripke says about objects, and about how objects of discourse are picked out in the actual world, we are mainly told what objects *are not*, and how they are *not* picked out. About a certain table T, Kripke says:

> [T]his table is wooden, brown, in the room, etc. It has all these properties and is not a thing without properties [that is, a "bare particular"], behind them; but it should not therefore be identified with the set, or 'bundle', of its properties. Don't ask: how can I identify this table in another possible world, except by its properties? I have the table in my hands, I can point to it, and when I ask whether *it* might have been in another room, I am talking, by definition, about *it*. I don't have to identify it after seeing it through a telescope. If I am talking about it, I am talking about *it* […]. Some properties of an object might be essential to it, in that it could not have failed to have them. But these properties are not used to identify the object in another possible world, for such an identification is not needed. Nor need the essential properties of an object be used to identify it in the actual world, if indeed it is identified in the actual world by means of its essential properties (I have up to now left this question open). (Kripke 1980: 52–3)

In the quote, it is clear that Kripke uses the demonstrative "it" to refer to what he takes to be the table itself, which cannot be identified with the set of physical properties of the table, or with any other set of its traits, such as its phenomenal properties. Then, we ask, what is the "it" that is pointed out? One tempting answer here is that "it" is an *essence*—in this case, the essence of the table. However, it seems as if this way of answering the question is *not* compatible with what Kripke says about Nixon. In that passage, Kripke explicitly says that the essential properties of a thing need not be used to identify the object in the

---

[8] This conjecture should be qualified in various ways, but I shall omit such qualifications here. For example, as the contemporary discussion of *Naming and Necessity* has shown, we sometimes think that we pick out one thing, where there are two or more things, and we sometimes make distinctions between kinds of things that have to be revised at a later stage.

actual world: we can speak of "it" without, I take it that he means, knowing these properties. Rather, as we saw above, Kripke thinks that we begin with objects that "we *have*, and can identify, in the actual world," and then *find out* their essences by asking modal questions (1980: 53). Then, what is Kripke talking about when he talks about "him" or "it" with reference to, respectively, Nixon and the table? The best answer seems to be: Nixon and the table. In other words, the objects Nixon and the table are picked out by picking out Nixon and the table. Unfortunately, this has little explanatory value.

6.4.2 Sameness and sortals

However, there seems to be one viable understanding of how Kripke "fixes" the objects of discourse: they are fixed by means of sortal identification. Consider the following examples:

> This table is composed by molecules. Might it not have been composed by molecules? […] [C]ould anything be this very object and not be composed by molecules? (1980: 49)

> If […] it is demanded that I describe each counterfactual situation purely qualitatively, then I can only ask whether *a table*, of such and such color, and so on, would have had certain properties; whether the table in question would be *this table*, table *T*, is indeed moot, since all reference to objects, as opposed to qualities, has disappeared. (1980: 52)

> Consider a particular pain, or other sensation, that you once had. Do you find it at all plausible that *that very sensation* could have existed without being a sensation, the way a certain inventor (Franklin) could have existed without being an inventor? (1980: 146)

Prima facie, these passages involve many sortal identifications, in terms of 'this table', 'sensation', and 'inventor'. (One could perhaps also argue that 'Franklin' involves sortal identification, in that it is a typical surname.) Let us assume that object identification, even in Kripke, proceeds by means of sortal identification. What is particular to sortal identification is that it involves *individuation*: objects are identified as *a man* or *a table*, and re-identified as *the same man* and *the same table*. This last comment also throws light on the notion of rigid designation. Above, we concluded that the question of sameness is prior to the question of whether a term is rigid or not. Why is 'Nixon' a rigid designator of Nixon? The standard answer is that 'Nixon' denotes *the same man* in all possible worlds. Notice that we do not take the fact that 'Nixon' does not denote *the same mereological object* in all possible worlds to count against the rigidity of 'Nixon'. This is because 'Nixon' is introduced as a name of *this man*, and not of *this mereological object*. Why is 'the watery stuff' not a rigid

designator? Because 'the watery stuff' does not denote *the same liquid* in all possible worlds (cf. Putnam's "same$_L$" in Putnam 1975), but only the liquid that plays the role of the actual watery stuff.

What seems to be evident from these examples and from those in *Naming and Necessity* is that we prefer to identify and individuate by means of certain sortals rather than others. There can be various reasons for this. One possible reason is that some sortals, and some concepts, capture the *nature* or *essence* of the thing better than others. However, this idea is problematic on the account developed here, because, strictly speaking, there is no "thing" prior to the sortal individuation. An alternative explanation is that among the different concepts we employ when making sortal identifications—"table," "sensation," "man," etc.—there are concepts that are more basic than others, or, alternatively, of which we have a firmer grasp than of others. For example, perhaps we have a firmer grasp of concepts such as "table" or "animal" than of "mental state" or "sensation"? A third position may be, roughly, that this is just how we speak. We preferably speak of tables and chairs and not of heaps of molecules, and in order to find an explanation of why we do so, one should turn to a study of psychological properties, not to properties of concepts or things. In the next, final, section of this chapter, I shall turn to a more extensive discussion of these issues.

In summary, the discussion thus far has suggested that metaphysical modality can be understood as dependent on sortal individuation. Given that things are identified in the actual world by means of sortal individuation, and that our main question, that of "sameness," is settled by means of sortal considerations, then the distinction between rigid and non-rigid terms is itself dependent on sortal individuation. But then, if the distinction between rigid and non-rigid terms is dependent on sortal individuation, the question of whether an a posteriori truth is (metaphysically) necessary will itself be dependent on sortal individuation, since necessary and contingent a posteriori truths are distinguished with respect to the question of whether they involve rigid designators or not. These ideas lie at the core of the conceptualist position outlined in section 6.5.3 below.

6.5 Essentialism, conceptualism, or skepticism?

In the following section, I will discuss a number of ways in which philosophers have viewed the relation between the identity, or essence,[9] of a thing, its modal

[9] Fine (1994) argues that the concept of essence is of central importance to the metaphysics of identity (1994: 3). This conjecture seems to be in line with Almog's views in his paper

properties (or the modal truths that apply to it), and what is conceivable with respect to it. When this discussion is completed, I shall be able to turn to my conclusions about the relation between conceivability and possibility in chapter 7, where I shall pick up central lines of thought from previous chapters, and conduct a unifying discussion.

6.5.1 Modal essentialism

Beginning with Kripke's view on the relation between the concepts of *essence*, *modality*, and *conceivability*, this view seems to be the following. As we indicated, the object of discourse is fixed without recourse to modal considerations. In chapter 5, we further argued that Kripke employs the *misdescription model of modal error*, according to which nothing is conceivable unless it is possible. In other words, according to our interpretation of Kripke, modal truths determine truths about conceivability.[10] Schematically, the structure of Kripke's arguments is the following.

Modal truths determine truths about conceivability:

(i) There is an object, identified, or "fixed," in actuality.

(ii) By counterfactual considerations (conceivability) about that object, we can come to know non-trivial modal truths about it.

(iii) Non-trivial modal truths about the object delineate its essential properties. For example, if Nixon could not have been inanimate, Nixon is, to be precise, *essentially* human.

(iv) Modal truths about the object *determine* conceivability truths about it (that is, what is and what is not conceivable about it). In cases where conceivability error occurs, this can be explained via the misdescription model of modal error (see section 5.4).

6.5.2 Non-modal essentialism

The above view can be contrasted with the view presented by Fine (1994) and Almog (1991). In some respects, one could say that the views of Fine and Almog are implicit in Kripke, in that Kripke holds that the object of discourse is "fixed" before modal considerations about the object. In other respects, Fine and Almog both make distinctions between essential facts on the one hand and

---

"The what and the how" (1991), which we shall consider below. Given that one can make sense of the notion of essence, we shall take it to be clear that essence and identity are closely related concepts.

[10] Our interpretation here is influenced by that in Almog 2002a.

modal facts on the other, distinctions that Kripke seems unable (or unwilling) to make, as his notion of essence is modally grounded (as is evident from (ii)–(iii) above).

In his paper "Essence and modality" (1994), Fine argues that the notion of essence should not be understood in terms of metaphysical modal truth. Rather, Fine believes that metaphysical modal truths are grounded in truths about essence. He writes:

> Indeed, it seems to me that far from viewing essence as a special case of metaphysical necessity, we should view metaphysical necessity as a special case of essence. For each class of objects, be they concepts or individuals or entities of some other kind, will give rise to its own domain of necessary truths, the truths which flow from the nature of the objects in question. (1994: 9)

More systematically, Fine agrees to the following thesis (1994: 4):

(15) If an object *x* has the property *F* essentially, then (if *x* exists) *x* has *F* necessarily.

However, he rejects the converse, that is:

(16) If an object *x* has the property *F* necessarily (whenever *x* exists), then *x* has *F* essentially.

Central to Fine's thesis that the notion of essence should not be understood in modal terms is his argument that the notion of essence and the modal notion of necessity are not extensionally equivalent. Fine takes essential truths to be statements that pertain to *what a thing is*, and modal truths to be statements that pertain to the way in which the world, with respect to the thing in question, could have been. With respect to the notion of essence, as we noted earlier, it is clear that Fine takes it to be closely related to the notion of *identity*. Fine writes:

> [O]ne of the central concerns of metaphysics is with the identity of things, with what they are. But the metaphysician is not interested in every property of the objects under consideration. In asking 'What is a person?', for example, he does not want to be told that every person has a deep desire to be loved, even if this is in fact the case.
>
> What then distinguishes the properties of interest to him? What is it about a property which makes it bear, in the metaphysically significant sense of the phrase, on what an object is? It is in answer to this question that appeal is naturally made to the concept of essence. (1994: 1)

Now, Fine argues that not all of the necessary properties of a thing pertain to what a thing is, or, equivalently, belong to its essence. Consider Socrates and the set with Socrates as its sole member, {Socrates}, or singleton Socrates. Fine notes that it is both necessary that singleton Socrates exists if Socrates exists,

and that Socrates belongs to singleton Socrates if he exists. In other words, the following statement is a necessary truth about Socrates:

(17) Socrates, if he exists, belongs to singleton Socrates.

However, Fine notes, there is no sense in saying that the property of "belonging to the singleton Socrates" pertains to *what Socrates is*, or to his essence: "There is nothing in the nature of a person […] which demands that he belongs to this or that set or which demands, given that the person exists, that there even be any sets" (1994: 5). Thus, (17) appears to be a *necessary* truth about Socrates, but it does not appear to be an *essential* truth about him. Therefore, the notion of essence and the modal notion of necessity do not coincide extensionally.

Almog (1991) provides a similar picture of the relation between essential truth and necessary truth. He employs the notion of "primal truths," which he takes to be "truths [that obtain] solely in virtue of what their subject is" (1991: 226). For example, the statement

(18) Nastassja Kinski is a human being

is, according to Almog, true in virtue of *what* Nastassja Kinski is (as opposed to in virtue of *how* she is) and is, as such, a primal truth (about Nastassja Kinski). It is clear from this that the notion of primal truths is similar to the notion of essence developed by Fine above; at least, it is clear that the question of whether a particular statement is a primal truth can be determined without recourse to modal thought experiments. Thus, it may be that Almog uses the notion of primal truth instead of the notion of essence in order to distinguish the notion he is developing from traditional modally grounded notions of essence.

The test that Almog devises for determining whether a statement is a primal truth consist in the following question: with respect to a subject *x* (such as Nastassja Kinski) and a property *F*, is it coherent to subtract the property *F* from *x*'s actual profile? (1991: 227. Almog calls this test the *coherent subtraction test*.) In essence, if *F* cannot coherently be subtracted from the actual profile of *x*, then *F* is a property that pertains to *what* an object is, and "*x* is *F*" is a primal truth (such as (18)). On the other hand, if *F* coherently can be subtracted from *x*'s profile, then *F* does not pertain to what *x* is, and "*x* is *F*" is not a primal truth. Here, the "actual profile" of an object is the set of properties that the object actually possesses, including properties that pertain to *what* the object is as well as properties pertaining to *how* it is. The goal of the coherent subtraction test is to determine what properties belong to the former category (the essential properties of *x*), and what properties that belong to the latter category (the contingent properties of *x*).

What does it mean, according to Almog, to say that it is *incoherent* to subtract a certain property from an object? In other words, what is the outcome of subtracting *F* from *x* if "*x* is *F*" is a primal truth? Almog writes:

> In subtracting a trait from the profile of […] a particular subject […] we must strike a delicate balance between (i) merely stripping away the subject of its predicates and (iii) preserving its integrity, viz., the sense in which it is still this subject from whose profile we subtract. (1991: 226)

Now, it is clear that like (18) above, "Nanga Prabat is a mountain" is a primal truth. Almog continues:

> But then what of one who claims that 'being a mountain' is subtractible [from Nanga Prabat] offering as his "proof" the coherence of a supposition, a "short story" if you will, in which N.P. itself bears proudly the predicate 'not a mountain'? […] Inevitably, the suspicion grows that our subject is not Nanga Prabat but merely an "arbitrary" entity, with no identity, a close kin to the philosopher's mysterious "bare particular." […] Gone is the sense that it is still about the real Nanga Prabat, not an invented philosophical fiction, that we suppose. (1991: 226–7)

From these passages, it is clear that Almog's idea is that incoherent subtractions from an object result in loss of the object itself, that is, that the discourse ceases to be about the (same) object. Notice that the central question of "sameness" that we discussed in connection with Kripke above is actualized at this point. We can also note that Almog's remarks echoes the skepticism towards "bare particulars" that we saw Kripke express above.

At this point, Almog proceeds to the distinction between primal (or essential) truth and necessary truth. First, he provides the following definitions: (i) a property *P* is a *condition of existence* for *x* iff it has to be actually true that, if *x* exists, *x* is *P*; (ii) a truth "*x* is *P*" is a primal truth iff *P* is a condition of existence (for *x*). Given these definitions, it is easy to see that, on the one hand, "is actually an actress" is not a condition of existence for Nastassja Kinski, but it is nevertheless necessarily true (that is, true in all possible worlds) that Nastassja Kinski is actually an actress. In other words, there are non-primal necessary truths. On the other hand, there are also primal non-necessary truths. For example, "I exist" is a primal truth, since it has to be actually true that, if I exist, I have the property of existence (see (i) above). However, it is not necessary that I exist.

In summary, Fine and Almog seem to hold the following view of the relation between essence (or "whatness"), possibility, and conceivability. First, two cautionary notes. To begin with, a more elaborate exposition of Fine's and Almog's views would show that they differ. However, these differences need

not, at least prima facie, prevent us from ascribing roughly the same view regarding the *relation* between essence, possibility, and conceivability to them. Rather, these differences would become crucial if we were to discuss the *origins* of (in particular) essential facts, and so on. Secondly, the relation between conceivability and essence that we outline below is not explicit in Fine, and only arguably explicit in Almog (1991; 2002a). However, here it is primarily the relation between *modality* and essence that we are concerned with.

Essential truths determine truths about modality and conceivability:

(i) There is an object, identified, or "fixed," in actuality.
(ii) By counteractual considerations (Almog) or (definitional) questions pertaining to what kind of object the object is (Fine), we can come to know essential truths about it.
(iii) Essential truths about the object *determine* modal and conceivability truths about it.

We note that (i) remains the same as in the conception regarding the relation between these concepts that we ascribed to Kripke. Below, we shall return to the question of how the object is picked out in the first place on Fine's and Almog's views.

6.5.3 Conceptualism

The third view of the relation between these concepts is the one we have outlined in sections 6.4 and 6.5 above. According to this view, modal metaphysical truths are derived from a priori conceptual (or "broadly logical") modal truths and physical facts. On this conception, there is only one type of modality involved in modal truths, regardless of whether they are a posteriori or a priori. In *From Metaphysics to Ethics* (1998), Jackson argues that the difference between a posteriori (or metaphysical) modal truths and a priori (or broadly logical) truths lies in the way we have *access* to them, not in *the sense* in which they are necessary:

> I think […] that it is a mistake to hold that the necessity possessed by 'Water = H_2O' is different from that possessed by 'Water = water', or, indeed, '2 + 2 = 4'. Just as Quine insists that numbers and tables exist in the very same sense, and that the difference between numbers existing and tables existing is a difference between numbers and tables, I think that we should insist that water's being H_2O and water's being water are necessary in the very same sense. The difference lies, not in the kind of necessity possessed, but rather where the labels 'a priori' and 'a posteriori' suggest it lies: in our epistemic access to the necessity they share. (1998: 69–70)

Our discussion in sections 6.3 and 6.4 suggests that the (single) type of modality involved in establishing metaphysical modal truths—by means of the model outlined—is conceptual, or broadly logical. We share this view with Tom Baldwin (2002). In his "Kantian modality," he distinguishes between three positions which he takes to be alternatives to what he calls "full-blooded realism," i.e. the view that possible worlds and possibilia exist in the same sense as the actual world exists, and all actual states of affairs exist. (This view is usually attributed to David Lewis. See Lewis 1973: 84–91.) These positions are *Aristotelian essentialism* (see section 1.2.2), *Kantianism*, which, according to Baldwin, involves the "view that modal concepts are categories, *a priori* concepts of understanding, whose warrant lies in the fact that they enter into the constitution of any possible conceptual scheme that provides for objective truth" (2002: 7), and finally, a *Humean* view according to which modal attributions are due to the irresistibility of certain judgments (see below, and also chapter 1).

Among these positions, Baldwin thinks that the Kantian position is the most attractive, and this is the position that he develops. In meeting what he takes to be the obvious objection to the Kantian position—that there are *de re*, a posteriori, necessities—Baldwin argues that one can account for the standard *de re* necessities while still describing the modality involved as conceptual. For example, consider the Kripkean thesis that all identical things are necessarily identical ((7) above). Baldwin argues that even if this thesis is applied to specific objects, such as Phosphorus and Hesperus ((9) above), and the resulting statement "necessarily, Phosphorus = Hesperus" is taken to be an ascription of *de re* necessity *about* the objects Phosphorus and Hesperus, one could argue that it is *qua objects* that it is necessary of Phosphorus and Hesperus that they are identical:

> [O]n the one hand, the concepts *object* and *identity* are sufficiently accommodating to encompass all the ways of describing or referring to [Phosphorus and] Hesperus; but, on the other hand, they still offer the prospect of an account of necessity that is rooted in the employment of concepts and not just in the brute structure of reality. (Baldwin 2002: 18)

We shall elaborate on Baldwin's views later on. For now, I shall take it that the conceptualist view on the relation between essence and modality is the following:

Broadly logical modal truths determine truths about possibility:

(i) An object is identified, or "fixed," in actuality by means of sortal individuation under a sortal *F*.

(ii) By counterfactual considerations pertaining to that object *qua F*-object, we can come to learn what is possible and necessary for it *qua F*-object.

(iii) There are no non-conceptual facts about essence. In so far as objects have "essences," they have essences *in virtue of* properties of the concepts under which they sort.

Regarding (i): in our discussion regarding Kripke above, our implicit idea was that object identification by means of sortal individuation carries with it certain conceptual truths tied to the use of the sortal in question. For example, if we identify Nixon as *this man*, we are implicitly committed to certain conceptual truths—whatever they are—that follow from identifying the object of discourse, the object under counterfactual consideration, as *a man*. An essential point of Baldwin's paper is that also our use of even more fundamental concepts that are not sortal concepts, such as *object* and *identity*, carries with it an implicit commitment to certain conceptual truths. With respect to the concept of identity, for example, Baldwin (2002: 18–19) mentions two such truths: the *necessity of identity* ((7) above), and the *necessity of origin*:

(19) $\forall x \forall y(x \text{ originates from } y \rightarrow \Box(x \text{ originates from } y))$.

We note that at least the former of these two theses is what we have described as a broadly logical modal truth pertaining to the identity of things. In other words, one way of describing the conceptualist outlook that we are developing is the following. There are certain a priori modal principles that govern the concept of identity. We have called these "broadly logical modal truths pertaining to the identity of things." Metaphysical modal truths are established by applying these principles to specific empirical things, which are picked out by means of sortal individuation—a conceptual activity which puts further restrictions, beyond those imposed by the use of concepts such as *identity* and *object*, on subsequent counterfactual considerations (see the Nixon example above).

Regarding (iii): a "conceptual" form of essentialism is suggested in the following passage from Baldwin:

> [I]t appears to me reasonable to hold that if there are good reasons for maintaining a type of essentialist claim at all, then these reasons will turn out to revolve around one or two concepts whose application brings with them the essentialist claim in question. Thus consider the 'necessity of origin' thesis: if there are good reasons to accept [(19)] then these reasons will concern the relationship between the concepts of a material object and its identity, on the one hand, and that of its origin, or occasion of coming into existence, on the other. Hence, if there is a necessity here, it is conceptual […]. (2002: 19)

In other words, Baldwin seems to argue that talk about essences can be intelligible, but that the modal truths that support essential facts are conceptual in nature.

In chapter 7, I shall return to the conceptualist position and discuss it further. One may ask how *conceivability* relates to modality and essence according to a conceptualist view. This is one of the questions I shall discuss in chapter 7.

6.5.4 Skepticism and psychologism

In the first chapter of this essay, I started out from the assumption that there is objective modal truth, and objectively true modal statements. In sections 1.2.1 and 3.1, I discussed Quine's and Hume's views on modality. According to Quine, things or states of affairs do not have modal properties independently of how we think of them or describe them (Quine 1961a: 155–6; see also Plantinga 1974: 22–4). Consider, for example, Paul Zwier, who is both a distinguished mathematician and a runner. Quine would argue that if we describe Zwier—pick him out—as a *mathematician*, then it will be correct to say about him that he is necessarily rational, but only contingently bi-pedal. On the other hand, if we describe him—pick him out—as a *runner*, it would be correct to say that he is necessarily bi-pedal but only contingently rational. Quine's view is similar to Hume's, on which necessary relations between concepts (ideas) do not obtain independently of thought, but are contributed by the mind (cf. Hochberg 1995: 499).

> [T]he necessity, which makes two times two equal to four, or three angles of a triangle equal to two right ones, lies only in the act of the understanding, by which we consider and compare these ideas […].[11]

In the following, I shall briefly consider a position due to D. H. Mellor, which is close to Quine's (and perhaps also Hume's). In his "Natural kinds" (1991 [1977]), Mellor argues against Putnam's conclusions from the Twin Earth thought experiment (Putnam 1975). In informal terms, Putnam's thought experiment appeal to our intuitions about what we would call 'water' in a determinate counterfactual situation. To begin with, on Putnam's (and Kripke's) theory, water is the stuff that is the same stuff as the actual stuff that comes out of the kitchen tap, and that stuff is H_2O. In turn, 'Water' is a rigid designator since it denotes the same stuff—H_2O—in each possible world (for details about the Twin Earth thought experiment, see sections 4.4.1 and 5.3.2 above). Putnam's conclusion is that, if we were to be transported to Twin Earth, in

[11] Hume 1978 [1739]: 166.

which the stuff that comes out of the taps is XYZ, we would *not* call that stuff 'Water'—although *the Twin Earthians* call it 'Water'—because 'Water' has the same extension on Twin Earth as it has here, namely H_2O. (However, after a while we might start to speak like the "natives" and refer to XYZ as 'water'.) In other words, Putnam's argument appeals to our intuitions about how we would use the name 'Water' in counterfactual situations, and Putnam assumes that we share his intuitions in so far as we would not call XYZ on Twin Earth 'Water'.

Mellor does not share Putnam's intuitions, which is evident from the following passage:

> I agree that 'water' had (tenselessly) the same extension in 1750 [when the molecular structure of water was unknown] as it had in 1950 [when the molecular structure of water had been discovered]; what I deny is that at either time the extension was different on Earth and on Twin Earth. […] It is indeed quite plain to my Fregean eye that in 1950, as in 1750, 'water' had the same extension on Twin Earth as it had here. There was water on both planets alike, and there still is. We simply discovered that not all water has the same microstructure; why should it? Because its microstructure is an essential property of water? Well, that is what's in question. (1991 [1977]: 126–7)

In other words, Mellor does not share the intuitions that Putnam's Twin Earth thought experiment is designed to elicit. Rather, where Putnam and Kripke reserve the term 'Water' for H_2O, Mellor uses it to name the *watery stuff* in each possible world (or, on each planet), and he implicitly questions why the former use of the terms should be preferred to the latter. He concludes his paper by noting that essentialists must suppose that "a property's 'importance' [is] a feature of the world independent of our beliefs and theories"—an idea that he explicitly denies (1991 [1977]: 134–5).

On a Quinean and/or Mellorian position, the view of the relation between essence, modality and conceivability could perhaps be described as follows:

Conceivability determines modal truths:

(i) Conceivability (in terms of how we would describe counterfactual scenarios) determines modal facts about possibility and necessity.

(ii) There are no objective facts about essence.

Regarding (i): consider, for example, Yablo's (1993) definition of what it means that something is conceivable:

(20) A statement S is conceivable for a conceiver *x* iff *x* can imagine a possible world which *x* takes to verify S.

A conceiver who shares Putnam's intuitions about the Twin Earth thought experiment will fail to imagine a world that he or she takes to be a world in

which water is something else than H_2O because he or she shares the intuition that something which is not H_2O is not water. Thus, no world *w* such that the conceiver takes *w* to verify "Water ≠ H_2O" will be imaginable, and "Water ≠ H_2O" will thus not be conceivable to him or her. On the other hand, consider a conceiver who, like Mellor, does *not* share Putnam's intuitions about the Twin Earth experiment. Such a conceiver will find a world that he or she takes to verify "Water ≠ H_2O" imaginable, since he or she will take, for example, Twin World to be such a world. (That is, the conceiver will take Twin World to be a world that contains water, and that water is XYZ in that world.) Since such a world thus *is* imaginable for the anti-Putnamian conceiver, "Water ≠ H_2O" will be conceivable for that conceiver on the definition (20). Now, since Mellor denies that properties—in this case, "being watery stuff" and "being H_2O"—are objectively "weighted" with respect to their importance to the kind in question, there are, on this view, prima facie no external facts that can be brought forth in support of either interpretation. Therefore, the conclusion that even facts about possibility and necessity ultimately depend on how *we* would describe a particular counterfactual scenario lies near at hand. In other words, we would say that conceivability—in terms of what we would and would not say about a certain counterfactual scenario—determines possibility (that is, the possibilities for that scenario).

First, could the world have been such that water was not H_2O? That is, is it *strictly* possible that water could have been something else than H_2O? The answer would be *yes* given that we are prepared to call something which is not H_2O 'Water', but the answer would be *no* given that we are not prepared to do so. According to this position, there is thus no question of "strict" possibility that goes beyond our linguistic dispositions.

6.6 Summary and discussion

In this chapter, I have considered different ways to assign content to the notion of *strict* possibility. The question I have pursued, in doing this, is the question of what facts are ultimately responsible for the truth of true modal statements—for truths about how the world could have been. In summary, given that we want an *objective* notion of strict possibility, two positions have presented themselves. On the first position, at least some modal statements are true in virtue of the nature of empirical things. On a second position, all true modal statements are ultimately true due to relations between concepts.

In pursuing the first line of thought, according to which a number of modal statements are true in virtue of the nature of empirical things, and are thus

metaphysically necessary, we argued that the notion of metaphysical modality ultimately depends on sortal individuation. This was argued in connection with the *modal essentialism* that we attributed to Kripke (1980). The idea is that sortal individuation carries with it certain conceptual truths tied to the use of the sortal in question. (For example, if we identify Nixon as *this man*, we are implicitly committed to certain conceptual truths—whatever they are—that follows from identifying the object of discourse, the object of subsequent counterfactual considerations, as *a man*.) In other words, there will be one or more *concepts*, tied to certain conceptual truths, which determine the possible outcomes of counterfactual considerations.

On a second view held by Fine and Almog, objects come with their essences in the *actual* world, and there is no need to turn to counterfactual considerations in order to find out what these essences are. Instead, we should be concerned with the question of what kind of object the object in question *actually* is, and what properties we can coherently assume that it is *actually* lacking. However, the "sameness" problem reappears as we consider Almog's response to the question of what happens when an *incoherent* subtraction of properties is made. His idea is that coherent and incoherent subtractions are distinguished by whether the object of discourse remains the *same* or not (see the quotes from Almog 1991 in section 6.5.2 above). Now, Almog rejects the idea that a "real definition" must be provided in order to answer the question of what an object is and, *a fortiori*, the question of when an object remains the same throughout some possible discourse about it:

> From the beginning of this paper […] I have presupposed, rather than argued for, my view that in articulating the whatness of many worldly subjects, as natural as Nastassja and the Nanga Prabat or as artificial as the Mona Lisa and the Notre Dame cathedral, "defining," better called "individuating," features need not be included. […] In a less presuppositional mode, I note that in response to 'what is Nastassja (Nanga Prabat, Mona Lisa)?' we do not despair and even appreciate 'a human being', 'a mountain', and 'a painting'. (1991: 233)

What Almog seems to be saying here is that the objects of discourse are given, and need not be individuated, before their "whatnesses" can be articulated. As Almog notes, the latter answers—'a human being', 'a mountain', and 'a painting'—to the relevant "what"-questions do not provide uniquely identifying descriptions of the relevant objects. However, they provide precisely the type of sortal identifications that we discussed in relation to Kripke. Similarly, Fine's central argument concerning Socrates and singleton Socrates seems to involve sortal identification. Fine argues that there is "nothing in the nature of *a person* […] which demands that he belongs to this or that set or which demands […]

that there even be any sets" (1994: 5). Therefore, there seems to be no reason why the arguments and ideas that we presented in that discussion cannot be brought to bear on the theories presented by Almog and Fine.

Given these results, the second position, according to which all true modal statements are ultimately true due to relations between concepts, seems the most plausible. However, this position is compatible with many views on the nature of conceptual truth. For example, in developing this position, we turned to Baldwin (2002). Baldwin's view on conceptual truth is anti-realistic (2002: 15): conceptual truths, or at least the conceptual truths that are constitutive of metaphysical modal truths are *not* held to be primitive, nor, I take it, objective features of reality (2002: 9). Instead, Baldwin views the conceptual truths that we have been concerned with as a form of obligations involved in the use of the concepts in question (concepts such as *identity* and *object*), and he argues that such obligations cannot be "disentangled" (2002: 15) from the relevant concepts themselves. However, as I see it, we have as yet not found any reasons to regard the conceptual modal principles we have been concerned with as anything else than objective features of reality, entering into, as Baldwin says, "the constitution of *any* possible conceptual scheme that provides for objective truth" (2002: 7, italics mine). In other words: a conceptualist view on modality, such as the one we have been developing throughout this chapter, is compatible with both conceptual realism and anti-realism. The main question is: are the conceptual truths that we have been concerned with truths about an objective conceptual reality, or are they true in virtue of something else? (If the latter, then they may be described as true in virtue of how we think, and the psychologistic position lies near at hand.) In chapter 7, we shall discuss if, and how, our conceptualist theory about modality can avoid psychologism and conventionalism.

CHAPTER 7

CONCLUSIONS AND DISCUSSION

7.1 Introduction

I shall begin this chapter by recapitulating the conclusions of chapters 1–5. In doing this, I shall argue that the best strategy for dealing with the standard counterexamples to the conceivability thesis is to reject the demand for universal applicability, and limit the scope of the thesis to a certain set of statements. Below, I attempt to delimit this set by means of epistemological considerations. Moreover, I try to accommodate this view with the account of possibility provided in chapter 6. In what follows, I introduce two additional theses concerning the relation between conceivability and possibility that, I argue, can be defended if one accepts the conceptual account of possibility presented in chapter 6 and the misdescription model of modal error outlined in chapter 5. I finally contrast the resulting modal epistemology with Descartes' and Arnauld's ideas, and with the ideas presented by Nagel (1998) regarding our current conceptual scheme.

In this thesis, I have been concerned with the questions if, and how, modal knowledge can be obtained. In chapter 1, I adopted an explanatory framework according to which *statements* are the primary bearers of modal properties. In chapter 2, I reviewed proposals on how we can obtain modal knowledge. I decided to focus on the thesis that *conceivability implies possibility*. In our terminology, this thesis is equivalent to the following:

(1) If a statement S is conceivable, then S is possible.

In chapter 3, I introduced the notion of *strict* possibility, on which S is possible just in case the world could have been such that S was true (as opposed to how the world *actually might be*). In chapter 4, I argued that this is the notion of possibility employed in many central arguments in metaphysics, including the mind-body arguments presented by Descartes, Kripke, and Chalmers. Thus, in these central applications of the conceivability thesis, facts about conceivability are taken to be facts about how the world could have been:

(2) If S is conceivable, then S is *strictly* possible.

In chapter 5, I considered various definitions of what it means that something is conceivable. Epistemological considerations suggested a number of desiderata that any notion of conceivability should satisfy. First, as we argued in chapters 3 and 4, the conceivability thesis should be interpreted in terms of strict possibility, that is, in terms of how the world could have been. Secondly, we should identify conceivability with *apparent* conceivability, that is, with that which, according to the conceiver, is what he or she conceives—or, with what appears to the conceiver that he or she conceives. Thirdly, the relevant form of conceivability must be attainable for us, or for conceivers with roughly the same cognitive capacities as ours. I argued that this is the notion of conceivability that Descartes is concerned with in the *Meditations*. To emphasize the connection with this notion, I shall in what follows refer to conceivers with roughly the same cognitive capacities as ours as "limited conceivers."

In sections 5.4–5.5, I argued that these desiderata are not jointly satisfiable. As the various counterexamples to the conceivability thesis have suggested, we should not expect the relation between conceivability and possibility to be such that conceivability *both* implies possibility, *and* that the relevant conceivability data are epistemically accessible to limited conceivers. In other words, we are presented with a dilemma: we have decided to interpret the conceivability thesis in terms of strict possibility, because this is the relevant interpretation in much philosophical argumentation. Now it seems that if we want to be certain that conceivability implies strict possibility, we must, epistemologically speaking, put constraints that are too demanding on what it takes for someone to conceive of something. On the other hand, if we tend to the epistemological problem and loosen the constraints, we can no longer be certain that conceivability implies strict possibility.

7.1.1 Conceivability is a global but fallible guide to possibility

We have noted that philosophers have responded in different ways to this dilemma. Yablo (1993) argues that the conceivability thesis should be interpreted as saying that conceivability is fallible evidence for possibility. On such an interpretation, the conceivability thesis can be defended even in the face of counterexamples. In section 4.2, we further saw Yablo (1990) argue that we should resist *gratuitous* skepticism towards conceivability intuitions, and take conceivability to be *prima facie evidence* for possibility. Here Yablo attempts to shift the burden of proof back to the skeptic, and challenges him to identify the point at which our inferences from conceivability to possibility have gone wrong. Similarly, Van Cleve holds that positive conceivability (where one can,

roughly, "see" or imagine that a statement S is true) constitutes prima facie evidence for the possibility of S. Van Cleve argues that if someone objects to this weak version of the conceivability thesis, he is not objecting to conceivability claims in general, he is "merely indulging in a quite general skepticism" (1983: 38). In summary, Yablo's and Van Cleve's thesis is roughly the following:

(3) If S is apparently conceivable for a limited conceiver *x*, then *x* is prima facie justified in believing that S is strictly possible.[1]

7.1.2 Ideal conceivability globally implies possibility

Chalmers (2002a), on the other hand, wants to maintain a strong link between conceivability and possibility. As we argued in section 5.5, this is done at the price of certain idealizations. With some terminological changes, Chalmers' proposal is:

(4) If S is ideally primarily conceivable for a conceiver *x*, then the primary proposition expressed by S is strictly possible.[2]

S is ideally conceivable for a conceiver *x* in the relevant sense if S is apparently conceivable for *x* and S's conceivability cannot be proved erroneous by a conceiver who possesses better reasoning capacities than *x*.[3] In defending his ideal notion of conceivability, Chalmers appeals to the concept of knowledge:

> The idea is that when prima facie conceivability falls short of ideal conceivability, then either the claim that the relevant tests are passed will be unjustified, or the justification will be defeatable by further reasoning. For ideal conceivability, one needs justification that cannot be rationally defeated. […] I note that the notion of undefeatability invoked here is also implicit in our concept of knowledge: it is

---

[1] Yablo (1990; 1993) is concerned with what is conceivable *for himself* (or, I take it, to anyone of us). It is thus natural to interpret him as being concerned with conceivability for limited beings, in the Cartesian sense.

[2] For Chalmers' conceivability theses, see section 5.3.3.

[3] Note that Chalmers' definition is compatible with the following claim, which concerns limited conceivers: *if* S *is ideally conceivable for a limited conceiver, then* S *is strictly possible*. S is conceivable for a limited conceiver *x* in the relevant sense precisely when *x* cannot be proved wrong by a conceiver with better reasoning capacities and/or more complete information. One could also consider "internal" definitions of ideal reasoning and information access, according to which S is ideally conceivable for any conceiver iff that conceiver, to the best of its reasoning capacities, cannot detect a contradiction in S even if provided with complete a posteriori information. Such definitions would amount to *local* idealizations—pertaining to the capacities of a specific conceiver—in contrast to *global* idealizations—relating to the capacities of all actual or merely possible conceivers.

> generally held that if one's justification for a belief that P is defeatable by better reasoning, then one does not know that P. So the notion of [ideal] conceivability is not obviously worse off than the concept of knowledge. (2002a: 148)

Similarly,

> One might worry that because the notion of ideal conceivability itself involves the notion of possibility (for example, in claims about what some possible being could conceive, or about what is defeatable) […] renders conceivability toothless as an epistemic guide to possibility, and so defeats modal rationalism. But this is not so: modal rationalism holds that modality is a priori *accessible*, and so invokes the notion of possibility in a precisely parallel manner. If ideal conceivability tracks possibility, then modal facts are rationally accessible, as required. (2002a: 172–3)

In other words, Chalmers argues that the notion of ideal conceivability is not worse off than the concept of knowledge, because the truth of both "*a* knows S" and "S is ideally conceivable for *a*" require that some external criterion is fulfilled. In order for a limited conceiver to *know* that S, he or she must have the proper attitude to S (he must believe that S, be justified in doing so, and his belief must be acquired in the right way), and S must be *true*. Here, the requirement that S is *true* is one external requirement. In order for a finite conceiver to *ideally conceive* that S, he must, again, have the proper attitude to S (he must, for example, imagine a scenario which he takes to verify S), and his conception must be *undefeatable*. Here, one external requirement is that of *undefeatability*.

In section 5.5, I argued, in response to Chalmers' strategy, that (4) might be useless for limited conceivers such as us. With respect to (4), how are we to know whether a certain statement is conceivable for a less limited conceiver? How can you ever know whether your own conception is ideal, if this ultimately is determined in terms of what a less limited conceiver can conceive? Chalmers seems to present us with the same situation as Arnauld did by his requirement for adequate knowledge (see section 4.2). In one sense, if we accept the ideal notion of conceivability employed in (4), we have slipped between the horns of the dilemma we presented above, but we have replaced the question of how we can know that something is possible with the perhaps equally difficult question of how we can know that something is "ideally" conceivable.

Thus, my conclusion in section 5.5 was that Chalmers' strategy does not solve the epistemological problems. Here I shall try to argue in a similar way, in analogue with Descartes' reply to Arnauld's requirement for adequate knowledge. Chalmers' point, by his appeal to the concept of knowledge, is that belief and apparent conceivability falls short of being knowledge and ideal

conceivability when some external requirement is not fulfilled (truth; undefeatability). Arnauld's objection to Descartes is similar: possibility cannot be inferred from conceivability unless an external requirement on our conceptions is fulfilled, namely that they should be adequate (or based on knowledge of all the essential properties of the thing in question).

Now, Descartes argues that in many cases, our conceptions are such that they fulfill Arnauld's requirement—"this could easily occur" (CSM 2, Rep 4: 155). This is something to which I believe Arnauld would agree. Likewise, I believe that most of us, and Chalmers himself, would concede that, in many cases, apparent conceivability *does not* fall short of being ideal. However, Descartes argues:

> in order for us to recognize a real distinction between two things it cannot be required that our knowledge of them be adequate if it is impossible for us to know that it is adequate. And since […] we can never know this, it follows that it is not necessary for our knowledge to be adequate. (Descartes, CSM 2, Rep 4: 155)

In other words, Descartes' reply to Arnauld is that in order for our conceivability intuitions to be valid, it cannot be required that our conceptions should be grounded on adequate knowledge, since it is impossible for us to know if and when our conceptions are so grounded. As Descartes argues, we cannot know this unless God grants us a "special revelation of the fact" that such knowledge obtains (CSM 2, Rep 4: 155). Without the latter form of knowledge, Descartes obviously holds, we cannot *know* if and when our conceivability intuitions are valid, and this is precisely what he requires. The same argument can be advanced against Chalmers. It may be true that the concept of ideal conceivability is not worse off than the concept of knowledge. Limited conceivers may have conceptions that are ideal, in the sense that they cannot be defeated, and their beliefs may constitute knowledge, in the sense that their beliefs are true (in the same sense as limited conceivers may possess adequate knowledge). However, as one is tempted to say with Descartes, in many cases, especially as regards modal questions, with respect to which beliefs cannot be verified, they cannot know that they possess such knowledge unless God grants them a special revelation of that fact.

Below, we too shall appeal to the concept of knowledge in order to support our position. However, as we shall see, our appeal is different from Chalmers'.

7.1.3 Conceivability locally implies possibility

Arnauld, Kripke, and van Inwagen employ a third strategy. Although they use different criteria in order to delimit the set of statements to which they believe

that the conceivability thesis can be applied, they at least agree that some such delimitation must be made. Thus:

> *Conceivability locally implies possibility*: with respect to a limited set *S* of statements, for each S in *S*, if S is apparently conceivable, even for a limited conceiver, then S is strictly possible.

In his exchange with Descartes, Arnauld argues that Descartes' conception of himself as possibly disembodied may be erroneous, but in his exchange with Leibniz, Arnauld claims that making a journey is not part of his essence, which is evident from his concept of himself:

> I am assured that as long as I think, I am myself. For I cannot think that I do not exist, nor exist so that I be not myself. But I can think that I shall or shall not take a particular journey, while remaining very much assured that neither one nor the other will prevent my being myself. So I remain very much assured that neither one nor the other is included in the individual concept of myself. (LAC: 33)

Similarly, van Inwagen argues that we often know modal propositions "that are of use to us in everyday life and in science and even in philosophy" (1998: 69), for example, that the legs and the top of this table might never have been joined to one another, but that we "do not and cannot know" that it is possible for there to be a perfect being, that it is possible that I exist and nothing material exists, or that it is possible that there exists a vast amount of suffering for which there is no explanation (1998: 68). Kripke (1980) argues that conceivability implies possibility (at least) with respect to certain subjective experiences that we cannot misrepresent to ourselves. In summary, there seems to be two ideas on how to delimit the set of statements *S* in the thesis that conceivability locally implies possibility. The first, due to Kripke, is that the conceivability thesis should be restricted to statements concerning states of affairs with which we are immediately acquainted through a privileged kind of representation. The second, due to Arnauld and van Inwagen, is that the conceivability thesis should be restricted to statements regarding a certain kind, or kinds, of *subject matter*.[4]

---

[4] As Gendler and Hawthorne (2002: 10) writes:

> Perhaps certain classes of propositions—abstract metaphysical ones, ones concerning necessary beings, ones that turn on actual empirical matters of fact, and so on—are illegitimate targets for conceivability–possibility arguments. Perhaps our conceiving faculty (whatever that turns out to be) is simply ill-suited to the task of providing reliable guidance concerning such realms. At the same time, it might be suggested, other topics are such that conceiving can deliver reliable modal verdicts concerning them. As long as we circumscribe subject-matter properly, conceivability will be a reliable guide to possibility.

7.2 The conceivability thesis as a local truth

Above, I have summarized the main conclusions of chapters 1–5. In what follows, I shall argue that the strategy employed in saying that conceivability locally implies possibility (that is, with respect to a certain set of statements) is to be preferred to the thesis that conceivability is global but fallible evidence for possibility (3). (Observe, however, that this is not to say that (3) is mistaken.) We have already objected to Chalmers' (4) as irrelevant to our concerns. However, I take Chalmers to be correct when he argues that Yablo and Van Cleve give up too soon, "settling for conceivability–possibility theses that are more attenuated than necessary" (2002a: 162). Yablo and Van Cleve both argue that the burden is on the skeptic to show that specific conceivability intuitions are invalid, and they further argue that gratuitous attribution of modal error should be rejected. I believe that they are correct in far as their claim is the following: one should not accept attributions of error *just because* the relevant judgments are based on conceivability intuitions. If the skeptic routinely rejects all judgments based on conceivability, then he or she is just, as Van Cleve argues, indulging in a general skepticism. On the other hand, it may be that someone who questions Van Cleve's or Yablo's *mind-body intuitions* is *not* indulging in a general skepticism. Rather, this person may doubt such intuitions in particular, because he sees a principled difference between modal claims about journeys and tables on the one hand, and claims about the nature of mind on the other. That is, this person sees a difference between ordinary events and states of affairs and abstract philosophical questions.

These remarks can be taken to show why the thesis that conceivability locally implies possibility is to be preferred to the thesis (3). Why should we only take conceivability to be a *fallible* but *global* guide to possibility when it seems possible to make principled distinctions between cases in which conceivability *implies* possibility and cases in which the relation between conceivability and possibility is uncertain? Is it not the task of the philosopher to find principled distinctions between cases in which conceivability implies possibility and cases in which it does not? As we recall, both van Inwagen and Kripke present such principled distinctions.

One can also argue that in treating the relation between conceivability and possibility as topic-neutral, the person who rests contended with (3) misses an important point raised by Descartes. According to Descartes, our ability to make correct modal judgments is conditioned by psychological limitations, and one should refrain from making modal judgments in cases where one does not "perceive the truth with sufficient clarity and distinctness" (CSM 2, Med 4: 41,

42). Further, Descartes' reply to Arnauld suggests that modal judgments are valid as long as "the power of knowing possessed by the intellect is adequate for the thing in question" (CSM 2, Rep 4: 155), but that "it is in the nature of a finite intellect to lack understanding of many things" (CSM 2, Med 4: 42). There is a clear sense in which one can understand Descartes as saying that the conceivability thesis is not topic-neutral, in that the power of knowing possessed by the intellect—in Descartes' case, the intellect of a created limited conceiver—may be adequate or inadequate *depending on* what thing the thing in question is.

In summary, we take Yablo and Van Cleve to be correct in arguing that one should not accept attributions of error just because one's judgments are based on conceivability intuitions. On the other hand, we do not agree that the best response to the various counterexamples to the conceivability thesis is to argue for a weaker relation between conceivability and possibility. Rather, we should aim at principled distinctions between cases in which conceivability *is* a mark of possibility and cases in which it is not.

I shall now provide my final argument for the thesis that conceivability locally implies possibility. I believe that Arnauld's positive modal knowledge claim has the same source as the claims that van Inwagen makes about statements concerning everyday objects. In so far as we know anything, that is, in so far as knowledge—modal or non-modal—is possible, we know that these statements are true. To doubt them would amount to a form of skepticism that not only undermines our *modal* thinking vis-à-vis the concepts involved in these statements, but also the very concepts themselves. In other words, if skeptical remarks are raised even at a very basic level of modal dialogue, the critic has moved on to a discussion broader than the discussion at hand—from modal skepticism to general skepticism. van Inwagen writes:

> The table could have been two feet to the left. Of course it could have. We *know* that. We know that the table's having been two feet to the left is not among the things that are intrinsically impossible" (1998: 75).

In assuming (as van Inwagen does) that certain basic modal judgments are trustworthy, as are certain judgments based on sensory experiences, one assumes that any criticism that can be directed against the former can also be directed against the latter, and that such criticism is not really directed against these claims *as such*, but against knowledge claims in general. Although Arnauld's claim about a particular journey actually involves philosophically problematical concepts, such as that of the self (the "I"), one can take his point to be the same. With respect to some questions, such as the question of whether the self (or the

"I") could exist without the body, Arnauld is hesitant to argue either way, and objects to Descartes' claims. However, even for such a concept as that of the self, Arnauld is certain that the concept of a *journey* is not substantially related to it (in the sense that I may entertain a thought of myself without, by positive exclusion, any thought about journeys). Thus, "I can exist without my body" and "I can exist, as myself, without having made a certain journey I in fact made," involves, I take it, in the first case, concepts between which the relations are unclear, and in the second case, concepts of which at least one can be clearly understood as unrelated to the other. If I am not allowed to infer, by means of my grasp of my concept of myself and my concept of journeys, that I would have been myself had I not made a certain journey, this does not only undermine my modal thinking with respect to the concepts involved, but also my grasp of the very concepts themselves. This is not to say that I have a firm grasp of the concept of myself (to the extent, for example, that I can know that it is possible for me to exist without a body). It is just to say that my grasp is firm enough to exclude journeys that I make (or not make) from having any substantial bearing on it.

Above, we saw that Chalmers appeals to the concept of knowledge in defense of an idealized notion of conceivability. Chalmers' claim is that the idealizations involved in his concept of ideal conceivability are similar to those involved in the concept of knowledge, and since we obviously find the concept of knowledge intelligible and useable, we should accept the notion of ideal conceivability too. Our appeal to the concept of knowledge in our arguments for the thesis that conceivability locally implies possibility has been different. In terms of knowledge and belief, we are arguing that, with respect to a certain set, or type, of statements, if our belief in these statements falls short of being knowledge, then we cannot know anything. In other words, some statements are such that only an extreme skeptic could doubt that they are true, and that our belief in them constitutes knowledge. As regards conceivability and possibility, our claim is that some statements—such as statements concerning everyday objects and events—are such that if their apparent conceivability is not also ideal, modal knowledge cannot be obtained, and our faculties for conceiving, as well as our modal concepts themselves, are fundamentally alien to modal reality.

7.3 A broader perspective

These arguments conclude my treatment of the matters discussed in chapters 1–5. However, these arguments do not extend to the mind-body intuitions presented by Kripke, whom we also took to endorse the thesis that

conceivability locally implies possibility. That is, these arguments apply to statements about tables and journeys, but not to statements concerning the nature of the relation between mind and body, or, to be exact, statements concerning the relation between phenomenal experiences and physical states or occurrences. I shall return to Kripke's account and the other mind-body arguments below, as I try to accommodate the conclusions from chapters 1–5 with the ideas in chapter 6, where I suggested a conceptual account of strict modality. In what follows, I shall introduce two further theses—besides the thesis that conceivability locally implies possibility—as regards the relation between conceivability and possibility. These theses can be defended if one accepts the conceptual account of possibility presented in chapter 6 and the misdescription model of modal error. In conclusion, I shall take these three theses to provide the outlines of a modal epistemology.

7.3.1 Ideal conceivability and strict possibility

Let us first consider Chalmers' theses about the relation between ideal conceivability and strict possibility. We have objected to idealized interpretations of the conceivability thesis due to epistemological considerations. However, this is not to say that the relevant conceivability theses are *false*. Chalmers takes the following thesis to be true:

(5) For all statements S, if S is ideally primarily conceivable, then S is primarily possible.

(5) says that if the primary proposition expressed by S is conceivably true on ideal reflection, then the primary proposition expressed by S is strictly possible, that is, true in at least one possible world. For example, "Water is not H_2O" is conceivable in the sense intended if a possible world in which the watery stuff is not H_2O is imaginable (via some kind of mental scenario). If this is imaginable, "Water is not H_2O" is also possible in the sense intended; there is at least one possible world in which the watery stuff is not H_2O. Chalmers further takes the following thesis to be plausible:

(6) For all statements S, if S is ideally secondarily conceivable, then S is secondarily possible.

(6) says that if the secondary proposition expressed by S is conceivably true on ideal reflection *informed by all the relevant facts about the actual world*, then the secondary proposition expressed by S is strictly possible. For example, "Water is not H_2O" is conceivable in the sense intended if a scenario in which water is not H_2O is imaginable. Since such a scenario is not imaginable in the

sense intended, there is no possible world in which the secondary proposition expressed by S is true.

Are (5) and (6) plausible? (We shall now set the *epistemological* arguments we have presented against idealized notions of conceivability to one side.) The central argument against (5) and (6) is, as I see it, based on the claim that there are what Chalmers refers to as *strong metaphysical necessities*. According to the two-dimensional analysis of statements that we have employed throughout the previous chapters, there is only one space of possible worlds, and the distinction between logical and metaphysical possibility (or, between primary and secondary possibility) is a distinction that applies on the statement level only. Roughly, on this view, each statement (of the relevant type) can be associated with two propositions. To repeat, a statement S is logically (or *primarily*) possible if the primary proposition expressed by S is true in at least one possible world, and a statement S is metaphysically (or *secondarily*) possible if the secondary proposition (that is, the proposition that the statement *actually* expresses) is true in at least one possible world. Now, Chalmers argues,

> There is a sense in which the truth of *statements* such as "Water is XYZ" is conceivable but not possible, but these examples never rule out the possibility of any conceivable *world*. They are merely instances in which such a world is misdescribed. (1996: 137)

"Water is XYZ" is conceivable in the sense that we can imagine a scenario in which the watery stuff is XYZ. In Chalmers' terms, this means that we can, via some kind of mental scenario, imagine a world in which the *primary* proposition expressed by "Water is XYZ" is true. However, when we then take this world to be a world in which *water* is XYZ, we have misdescribed it. What we have imagined, via the scenario, is not a world in which water is XYZ, but a world in which watery stuff is XYZ.

Chalmers emphasizes that these modal mistakes never rule out the possibility of any conceivable world. (This is a claim that I shall return to in the next section.) However, according to the view that there are strong metaphysical necessities, Chalmers argues,

> there are worlds that are entirely conceivable, even according to the strongest strictures on conceivability, but which are not possible at all. […] On this position, "zombie world" may correctly describe the world that we are conceiving, *even according to a secondary intension*. It is just that the world is not metaphysically possible. (1996: 137, my emphasis)

In other words, according to the view that there are strong metaphysical necessities, there are worlds that are conceivable even on the ideal notions employed in (5) and (6), that are not misdescribed, but which nevertheless are

not metaphysically possible. Obviously, the view that there are strong metaphysical necessities poses many problems, on many levels, for any version of the conceivability thesis.

What would it take for the view that there are strong metaphysical necessities to be true? According to this view, we can (via a scenario) imagine a world *w* that we take to verify a statement S (with respect to either S's primary or secondary intension), such that *w* does verify S (with respect to the same intension), but where *w* nevertheless is metaphysically impossible. Since this is taken to hold even with respect to *secondary* conceivability, a statement S can, on the view that there are strong metaphysical necessities, be conceivably true on ideal reflection *informed by the relevant non-modal empirical facts* without being metaphysically possible. Thus, the metaphysical impossibility of S cannot, even in principle, be detected by means of reflection on the concepts employed in S informed by sufficient non-modal information. This must mean that strong metaphysical truths are neither conceptual nor empirical truths.

However, according to the theory of possibility we have suggested in chapter 6, all modal truths are ultimately grounded in conceptual truths. Chalmers argues that the view that there are strong metaphysical necessities leads to a proliferation of modalities, and that such modalities will be brute and inexplicable (1999a: 483). The worry is the following: what is there, once we have assumed that things themselves have one type of modal properties, that prevents us from assuming that things have countless other modal properties (some of which are "brute and inexplicable")? Since strong metaphysical necessities are by their nature intractable, there is no real good argument for or against them once the thesis that things themselves have modal properties has been accepted. Our theory, on the contrary, does not assume that things themselves have modal properties (in the literal sense). In other words, the assumption that in the end might lead to a proliferation of modalities is not made.

Given that there are no strong metaphysical necessities, it is reasonable to argue that (5) and (6) are true. Since the relevant form of conceivability is ideal in terms of conceptual grasp only, or in terms of both conceptual grasp and a posteriori information, what has been conceived cannot be subjected to a misdescription style interpretation. However, there are even more compelling reasons for accepting (5) and (6), or even stronger theses, given a conceptual theory of modality. According to some of the positions we considered in chapter 6, truths about essences determine truths about conceivability and possibility. This makes for a tripartition of truths: truths about essences; truths

about possibility that supervene on truths about essence, and truths about conceivability, that are conceptual. A conceptual theory of modality provides no reason why these notions should be treated as separate notions. On this theory, conceivability proceeds by way of conceptual analysis informed by empirical information, and modal truths hold in virtue of relations between concepts and empirical truths. An ideal *conception* must thus be a conception that proceeds by way of a sufficient grasp of the relevant concepts and sufficient empirical information, where the correct broadly logical principles are applied to these concepts and this information (for such broad logical principles, see chapter 6). However, where we take these criteria to dictate what it means that something is ideally *conceivable*, chapter 6 argues that they also dictate what it means that something is strictly *possible*. On a conceptually grounded theory of modality, there are no facts besides facts about concepts and non-modal empirical facts, that affect the possibility of a statement without affecting the (ideal) conceivability of the same statement, or vice versa. Thus:

> *Ideal conceivability equals strict possibility*: S is ideally conceivable iff S is strictly possible.

Is this to say that ideal conceivability and strict possibility are the *same property* of statements? If we say that they are, we seem to render our position inconsistent, since in section 5.2.2 we have argued that ideal conceivability is a relational property of statements, which implies the existence of a (possible) conceiver, whereas we have assumed throughout that strict possibility is an intrinsic property of statements. Ideal conceivability is an epistemological notion, and we define what it means that a statement is ideally conceivable by means of an epistemological claim, according to which the relevant property is relational. (It is in part characterized in terms of the abilities of a possible reasoner.) However, the notion of strict possibility is a *metaphysical* notion, and what it means that a statement is strictly possible is characterized by means of a *metaphysical* claim, according to which the relevant property is an intrinsic property, that obtains independently of the existence of (possible) conceivers. Even if our theory proposes a very strong connection between ideal conceivability and strict possibility, it seems that we cannot argue that ideal conceivability and strict possibility are the same property of statements. However, we would do better in modifying our conceptual framework regarding these matters. As noted, ideal conception proceeds by way of sufficient grasp of conceptual relations and empirical matters, and strict possibilities obtain in virtue of conceptual relations (applied to empirical matters). In other words, both the notion of ideal conceivability and the notion of strict possibility are

themselves (equally) grounded in conceptual relations as applied to empirical circumstances. We can thus say that it is the *notions* of ideal conceivability and strict possibility that can be described as being relational or not—whereas the (one and the same) underlying property cannot.

7.3.2 Imaginability and impossibility

Finally, I shall address a thesis that seems compatible with the parts of Chalmers' account that we have accepted. In his paper, Chalmers (2002a) provides the following analysis of what possible sources of modal error there are, given that one accepts (5) and the various distinctions he makes between different forms of conceivability (see chapter 5). First, he notes, modal errors will stem from either the difference between prima facie and ideal conceivability (that is, where one mistakenly takes prima facie conceivability to be ideal) or when one confuses primary and secondary conceivability (that is, where one mistakenly takes an imagined world where the watery stuff is XYZ to be a world that verifies "Water is XYZ"). Chalmers argues as follows (2002a: 173):

> (i) Prima facie negative conceivability judgments can go wrong in cases where a 'deep' a priori contradiction is not revealed by prima facie reasoning.
>
> (ii) Prima facie positive conceivability judgments can go wrong when (a) an imagined situation that is taken to verify S does not in fact verify S, upon rational reflection; or when (b) an imagined situation is not coherently imagined, because of the failure to notice a deep contradiction, or because of the inability to fill in the crucial details.
>
> (iii) Primary conceivability judgments can go wrong if a subject mistakenly expects them to be a guide to secondary conceivability.
>
> (iv) Prima facie secondary conceivability can go wrong as a guide to secondary possibility when a subject is misinformed about relevant nonmodal empirical facts, or perhaps when an incautious subject is merely ignorant of those facts.

Does Chalmers' typology of modal error imply that we can imagine the impossible? With respect to (iii) and (iv), the modal errors dealt with here are covered by the misdescription model. One example of a type (iii) error is when a conceiver imagines liquid stuff being XYZ, and takes this imaginability to imply that water could have been XYZ. An example of a type (iv) error is when a conceiver imagines *that* (the morning star, ostensibly identified in the morning) being distinct from *that* (the evening star, ostensibly identified in the evening), without knowing that both things pointed out are identical to Venus.[5]

---

[5] This case is dealt with by Kripke, in a characteristically misdescription model style. See Kripke 1980: 102–4.

Regarding (i), negative conceivability judgments are known to be prone to modal error. (As we explained in section 5.3.1, a statement S is *negatively conceivable* iff S is not *ruled out* by what one knows or by what one believes.) What is notable here is that the negative account of conceivability does not make any reference to the faculty of imagination. In fact, negative conceivability has—at least not in the forms we have studied—nothing to do with imaginability at all. (This fits well with Chalmers' account, which explicitly associates *positive* conceivability with activities of the imagination.) In other words, no matter what modal mistakes we might make employing a *negative* notion of conceivability, they do not say anything about our ability to imagine the impossible.

Left to be discussed is (ii). Do the types of modal mistakes that Chalmers discusses under (ii) imply that we can imagine the impossible? Here, the mistakes discussed pertain to the scenario imagined: the scenario imagined may be such that it in fact does not verify a certain statement S that the conceiver takes the scenario to verify, or it might be such that it is not coherently imagined. Cases like these are discussed by Yablo (1993). As we indicated in section 5.4.1, Yablo suggests that his imagining of a computer producing a proof hailed as the refutation of Goldbach's conjecture falls short of being an imagining of a scenario in which Goldbach's conjecture is *false*, because his imagining may equally well be described as an imagining of a scenario in which Goldbach's conjecture is *true*, and where the proof produced by the computer is invalid. What does it take to imagine a scenario in which Goldbach's conjecture is false? Yablo writes (where "GC" is shorthand for Goldbach's conjecture):

> [L]et me describe some scenarios I clearly *can* imagine and then show how imagining *these* falls short of imagining that not-GC. For instance, I find it easy to imagine a computer printing out some unspecified number *n*, and this being hailed on all sides as an authentic counterexample. Why wouldn't this be a case of imagining that not-GC? Because it suffices for the veridicality of *this* imagining for the following to be possible: GC has no counterexamples, but the computer produces a number *n* widely though erroneously hailed as a counterexample. […] Maybe I do better to imagine the computer producing something widely acknowledged as a *proof* that *n* is a counterexample. But again, the proof can help me enjoy the appearance that possibly not-GC if it is imagined to be *correct* […] Am I to imagine it set out in convincing detail? But if the detail is only *imagined* to be convincing, it does nothing to increase my *actual* confidence in the proof's correctness. Am I to imagine the proof set out in *actually* convincing detail? If I could, I would call a press conference to announce my refutation of Goldbach's conjecture! (1993: 32)

It seems that Yablo's argument is that all but the last imagined scenario are compatible with Goldbach's conjecture being *true*. Suppose that Goldbach's conjecture is in fact true. It is clear that for each of the imagined scenarios except for the last one, I take these scenarios to be scenarios in which Goldbach's conjecture is *false*. On the assumption that Goldbach's conjecture is in fact *true*, does this entail that I have imagined something impossible? No, it does not: what I have *imagined* can, as Yablo points out, be expanded into a bigger scenario that, on the assumption that Goldbach's conjecture is in fact true, is *possible*—namely a scenario in which the proof that the computer produces is invalid but nevertheless hailed as correct. In other words, one can, in some sense, *conceive* the impossible, in that one can *take* the imagined scenario to be a scenario in which Goldbach's conjecture is false. However, this does not entail that I can *imagine* the impossible, since the scenario *actually imagined* can also be taken to be a scenario in which a computer produces an invalid "proof" of the negation of Goldbach's conjecture.

As I see it, these observations cover all the variants of modal error that Chalmers considers under (ii). First, the Goldbach example demonstrates how a scenario is taken to verify a statement S ("Goldbach's conjecture is false"), but where this scenario does, in fact, not verify S. Second, on the assumption that Goldbach's conjecture is *true*, it is an example of where my prima facie conception fails to reveal a "deep" contradiction (without implying that what actually has been imagined is contradictory). Third, it is certainly an example of where I am unable to fill in the crucial details. As Yablo says, if it is required that the crucial details be filled in "in *actually* convincing detail," I would, on the assumption that Goldbach's conjecture is *true*, be unable to imagine it as false.

From this, the step to a negative claim concerning the imaginability of the impossible is not far. Chalmers takes it that the examples of modal error we have discussed never rule out the possibility of any conceivable, or imaginable, *world*.

> There is a sense in which the truth of *statements* such as "Water is XYZ" is conceivable but not possible, but this never rules out the possibility of any imaginable *world*. They are merely instances in which such a world is misdescribed. (1996: 137)

The reason why Chalmers takes the modal errors we have discussed not to rule out the possibility of any imaginable world might be because he takes the set of imaginable worlds to *be* the set of possible worlds. Let us assume that the set of possible worlds is identical to the set of imaginable worlds. Taken together with

the misdescription model, the thesis that the set of imaginable worlds is identical to the set of possible worlds can be seen as promoting the following view of the relation between apparent conceivability, what is actually imagined, and possibility. Any imaginable scenario can always be expanded so as to correspond to at least one possible world. For example, the scenario I imagine that I take to verify "Water is XYZ" can be expanded so as to correspond to a world in which watery stuff is XYZ. In other words, on the level of *what I take myself* to imagine, even impossible statements, such as "Water is XYZ" are conceivable. However, on the level of what I *actually* imagine, what I imagine is always something that can be filled out in ways that make it possible. In other words, what I actually imagine is always something manifestly possible, that I misdescribe, or at least something that is fragmented enough *not* to make it manifestly *impossible*. One could say that *conceivability always implies some possibility*. This means the following:

> *The unimaginability of the impossible*: the details of that which is *actually* imagined can always be filled in so as to correspond to at least one possible world.

The corollary thesis about the relation between conceivability and impossibility is the following:

> *The conceivability of the impossible*: in assessing an imagined scenario, we may misdescribe it as a scenario that verifies some impossible statement S.

One of the comments we made in section 5.4.1 can also be taken to speak in favor of the thesis that we cannot imagine the impossible. We argued that our representations are seldom detailed enough to admit of only one interpretation, and that one cannot simply *will* one's representation—the imagined scenario—to be a representation of a specific state of affairs. What the above thesis says is that our representations are *never* detailed enough to exclude each interpretation according to which the object of imagination—that is, what has *actually* been imagined—is some *possible* states of affairs.

7.4 Outlines of a modal epistemology

In section 7.3, I have tried to broaden and develop the view of the relation between conceivability and possibility from chapters 1–5. The arguments in section 7.3 also rely on the conceptual theory of modality suggested in chapter 6. In summary, what is the resulting modal epistemology? What is the view of

the relation between conceivability and possibility that has been developed in this thesis?

According to this view, the ultimate basis of modal truths is broadly logical modal truths that explicate necessary relations between concepts. In some cases, modal truths are established by applying such broadly logical modal truths to empirical truths. This view of modality does not ascribe modal properties to things themselves, in the sense that things themselves have modal properties independently of conceptual relations. Therefore, there is no reason to believe that there are so called "strong metaphysical necessities," that is, modal truths that obtain independently of, on the one hand, the semantics of proper names and kind terms, and, on the other, independently of conceptual relations. Since our theory holds that *all* modal truths ultimately have their grounds in conceptual relations, there are no "brute and inexplicable" modal truths—*given* that the relevant concepts are accessible to us. The question of whether the relevant concepts are in each case accessible to us will to some extent be discussed below.

One consequence of this view is that since both conceivability and possibility are determined by conceptual relations, and in some cases, empirical facts, *ideal* conceivability can be seen as equivalent to strict possibility. In cases where a posteriori information is relevant (for example, with respect to ideal secondary conceivability), it will be enough that the conceiver is in possession of sufficient non-modal empirical information. In section 5.3.3, we noted that, when introducing the notion of ideal secondary conceivability, Chalmers thinks that "it is probably best to restrict the empirical knowledge in question to nonmodal knowledge" (2002a: 159) in order to avoid trivializing the relation between ideal secondary conceivability and possibility. In some sense, Chalmers is here implying that (or he at least leaves it open whether) there is *empirical* modal information. According to our theory, all (strictly) empirical information is non-modal, since empirical things do not have modal properties independently of conceptual relations. All *modal* "information" ultimately derives from a priori, broadly logical principles, and therefore we can simply say that S is ideally secondarily conceivable iff S is conceivable on ideal rational reflection informed by the relevant empirical information.

According to our theory, there is no reason why we should even *distinguish* between ideal conceivability and strict possibility. Both are equally grounded in conceptual truths as applied to non-modal physical facts. Strictly speaking, we say that whereas the notions of ideal conceivability and strict possibility can be differently defined, the singular underlying property of statements cannot.

Thus, our theory proposes the strongest possible link between conceivability and possibility with respect to ideal conceivability. However, one of the central theses of the theory we have proposed is that in order for the conceivability thesis to be useful, a statement S should not be described as conceivable unless it is conceivable for limited conceivers such as ourselves. This methodological imperative is put forward in conjunction with the theses about our conceptual and imaginative abilities that I have presented above, which (in a slightly different order than they have been introduced) are the following:

(i) What we *actually* imagine can always be expanded, or filled in, so as to correspond to at least one possible world. In other words, we cannot directly imagine something manifestly impossible. To take it the other way around: suppose that you in fact could imagine manifestly impossible scenarios, and that $\mathcal{S}$ is an imagined scenario which is manifestly impossible. Of course, such a scenario cannot be expanded into *any* possible world. If a scenario can be expanded into at least one possible world, then it is not manifestly impossible.

(ii) A modal mistake consists in *misdescribing* the content of imagination. We can thus, in one sense, *conceive* the impossible. When assessing an imagined scenario, we may *describe* it as impossible, and had the way we describe it been the *only* possible description of the contents of our imagination, it would *be* an impossible scenario. However, the examples we have considered suggest that imagination is always indeterminate. Recall Yablo's "striped tiger" example (see section 2.6). When I imagine that there is a tiger in the hallway, I do not determinately imagine it as having *n* stripes, but rather imagine it as determinate, that is, as for some *n*, having *n* stripes.[6]

(iii) For a certain set of statements, conceivability for limited conceivers implies strict possibility. To doubt that this relation obtains amounts to a form of skepticism that not only undermines our *modal* thinking by means of the concepts involved in these statements, but also the very concepts themselves. It is difficult to circumscribe this set of statements, because even with respect to philosophically laden concepts, such as that of the self, one can possess modal

---

[6] In some cases, it may be more proper to speak of *overdescription* than of misdescription. When I, for example, imagine a man on a beach standing on all four, I am, in a (rough) sense, overdescribing the content of my imagination if I describe him as crawling ashore (given that I do not imagine him moving). That is, the actual content of my imagination could as well be taken to represent a man crawling backwards to sea. (At least, I could use the *same image* in both cases.) However, the term *misdescription* is properly used for some cases *given* that one accepts the thesis that one cannot imagine something manifestly impossible. For example, if I describe my imagined scenario as a scenario in which water is XYZ, I am, *given* that one cannot imagine something impossible, *misdescribing* my imagined scenario.

knowledge. Maybe "I can exist without my body" does not belong to the relevant set of statements, but "I can exist, as myself, without having made a certain journey I in fact made" surely does. To clarify the point I want to make: if I cannot infer, from the conceivability of the table being two feet to the left, or of not having made a certain journey that I in fact made, that these things are possible, then I do not, in any useful sense, *possess* the concepts of tables, distances, journeys, and myself. Our grasp of some concepts may be imperfect, but it is not that imperfect.

Now, the question of conceivability is not just a matter of assessing the content of imagination, it is also a matter of assessing conceptual relations, quite apart from any workings of the imagination. These last remarks about concepts bring us to one final issue, which we left unresolved at the end of chapter 6. The question was: is our conceptual theory of modality committed to some form of psychologism or conventionalism? I believe that the answer is both yes and no, but more importantly—no. Let us begin by considering the sense in which the answer to this question is "yes."

I believe that some of our concepts are concepts of our own making. In one sense, in accepting that some concepts, such as those of cups and tables, are concepts *of our own making*, we are conceding to the conventionalist that some parts of our current conceptual scheme are dependent on us, and that the modal properties of such concepts, such as those of cups and tables are dependent on how we speak (*since* the concepts of tables and cups are concepts of our own making). How the world could have been, with respect to tables, is ultimately dependent on the concept of table. However, this is a concept of our own making, and we could have employed a conceptual scheme that did not include the concept of table at all. When Kripke argues that a certain table could not, in the strict sense, have been made of ice, he is correct. Something could not have been that *table* and been made of ice. However, on our view, this just relates to what is involved in the concept of table.

What is more important, I believe, is that some concepts are *not* concepts of our own making. As I suggested at the end of chapter 6, I believe that some concepts, such as logical and mathematical concepts, are fundamental to all possible conceptual schemes. I further believe that these concepts do not exclusively belong to the logical and mathematical domains, but also include concepts such as the concept of *consciousness*, the concept of a *conscious being*, the (general) concept everyone has of himself as a conscious being, and perhaps that of a *phenomenal experience*. What would a conceptual realism with respect to this latter group of concepts entail with respect to the mind-body problem? I

shall address this question by considering, very briefly, the views of two philosophers.

In his defense of non-reductive physicalism, Thomas Nagel argues as follows:

> […] I believe that there is a necessary connection in both directions between the physical and the mental, but that it cannot be discovered *a priori*. Opinion is strongly divided on the credibility of some kind of functionalist reductionism, and I won't go through my reasons for being on the antireductionist side of that debate. Despite significant attempts by a number of philosophers to describe the functional manifestations of conscious mental states, I continue to believe that no purely functionalist characterization of a system entails—simply in virtue of our mental concepts—that the system is conscious. (1998: 337)

Nagel argues that the ultimate hindrance preventing us from recognizing a particular system as conscious lies in our current conceptual scheme, and in our current concept of consciousness. He continues:

> […] I believe it is not irrational to hope that some day, long after we are all dead, people will be able to observe the operation of the brain and say, with true understanding, 'That's what the experience of tasting chocolate looks like from the outside.'
>
> Of course we already know what it looks like from far enough outside: the subject taking the first reverent mouthful of a hot fudge Sunday, closing his eyes in rapture, and saying 'Yum.' But I have in mind some view or representation of the squishy brain itself, which in the light of our understanding we will be able to *see* as tasting chocolate. While this is at the moment inconceivable, I think that it is what we would have to grasp what must be the truth of these matters. My reading of the situation is that our inability to come up with an intelligible conception of the mind and body is a sign of the inadequacy of our present concepts, and that some development is needed. (Nagel 1998: 338)

Nagel argues that if our conceptual scheme underwent some suitable revolution, we would with true understanding be able to represent states of the brain itself *as* the tasting of chocolate. In other words, had the proper conceptual revolution come to pass, our concept of conscious or phenomenal experiences would be imbued with a physical aspect. Our present inability to come up with an intelligible conception of the (necessary) relation between consciousness, conscious or phenomenal experiences, is a sign of the inadequacy of our present concepts, and is "due to the limitations of our understanding" (1998: 342).

Here, I believe that Nagel is wrong. As I argued, there is a set of concepts that is not of our own making, concepts that are *constitutive* of each conceptual scheme. It is not unreasonable to hold that the concept of consciousness is such a concept; at least, it is difficult to conceive of a conceptual scheme without this concept. I believe that even if such a conceptual revolution as Nagel hopes for will come to pass, if that conceptual revolution imbues our concept of

consciousness and conscious experiences with a physical aspect, we have not *developed* our *current* concept of consciousness, nor our concept of a conscious or phenomenological experience, but rather abandoned it altogether. At this point, one might want to remind oneself of the following passage in Jackson:

> It is always open to us to stipulate the situations covered by the various descriptive terms, in which case we address subjects of our stipulation rather than the subjects the titles of our books and papers might naturally lead others to expect us to be addressing. (1998: 42)

I believe that the history of the mind-body problem contains some evidence that the concept of consciousness and related concepts are irreducible, and fundamental to not only our way of viewing the world, but to any view of the world. Towards the end of chapter 4, I suggested that there is a sense in which Chalmers' reply to his critics is similar to Descartes' reply to Arnauld. Descartes' claim that although he may have many properties of which he is not yet aware, he is certain that he could have been created by God without having the properties of which he is *unaware*, is similar to Chalmers' claim that the logical possibility of a world wherein the (directly perceptible) "what it is like" properties are lacking—disregarding what *other* properties he (Chalmers) might possess—suffices to support dualism. In a very broad sense, both Descartes and Chalmers can be interpreted as saying that *whatever* truths that will be revealed about the relation between mind and body (or phenomenal experiences and body), it will still be possible to positively understand mind, or the relevant aspects of mind or consciousness, as distinct from body. According to our theory, that this possibility always remains open must ultimately be explained in terms of the relation between the relevant concepts. Furthermore, since what is possible is determined by conceptual relations, if *one considers the above concepts in isolation* (mind, body, consciousness), consciousness without body is possible. However, this possibility may be in conflict with the relation between *other* concepts, as we shall see in the following reading of Almog (2002a; 2002b).

In a number of works, Almog (2002a; 2002b) has argued that mind and phenomenal experiences on the one hand, and body and physical states on the other, are essentially connected but are nevertheless different *kinds* of things, and thus numerically distinct (that is, mind and phenomenal experiences are one kind of things, and body and physical states are another kind of things). That the first two are *essentially connected* with the latter two (mind with body, pain with C-fibers firing) means that they are connected by their very nature: by the very nature of mind, it is connected with body, and by its very nature, pain is

connected with a certain physical state (C-fibers firing). In Almog, these essential connections give rise to the *modal inseparability* of mind and body, and of pain and C-fibers firing: there is no possible world in which the one (mind; pain) exist without the other (body; C-fibers firing). Why are mind and body, or to be more specific, say, Descartes' mind and Descartes' body, essentially and modally inseparable? Because they make up a third entity, the *human* Descartes: they are the mind and body *of* a human being, which is a third type of thing, and for which other modal laws hold.

Our view is the following. There are three concepts involved here: that of mind, that of body, and that of human beings (which we shall take to be subordinate to the broader concept of *a being*). Considered in isolation, the first and the second are independent: mind without body seems possible. However, this possibility does not obtain with respect to *Descartes'* mind, and *Descartes'* body. In this case, the concept of a *human being* will be in conflict with the possibility of mind without body, that is, there is no possibility of *Descartes'* mind existing without *Descartes'* body. In addition, it may be that the possibility of mind or consciousness without body is in conflict with even more *basic* concepts, which we also take to be fundamental to each conceptual scheme, such as that of *a being*, that of *existence*, and so on. Could things really be this simple? Of course they could. However, if you are a conceptual realist, you may ask to what extent the difference between this view and a view such as Almog's is a real difference, as opposed to a difference in words.

Our answer is that there is a difference between the two views, as the latter argues that entities have conceptually independent essences—in other words, that an entity is what it is, no matter what concepts there are, or what concepts the entity sorts under. Our view is rather that an entity has a set of properties, none of which is metaphysically more significant than the others, independently of the concepts under which the entity sorts.

BIBLIOGRAPHY

In this thesis, references to Descartes, Arnauld, and Leibniz are abbreviated as follows: 'CSM 2' for Cottingham, J., R. Stoothoff, and D. Murdoch (eds.), 1984; 'CSMK' for Cottingham, J., R. Stoothoff, D. Murdoch, and A. Kenny (eds.), 1991; 'LAC' for Mason, H. T. (ed.), 1985 [1967]. In references to CSM 2, '4 Obj.' refers to the *Fourth set of objections* by Arnauld (pp. 138–53), and '4 Rep.' refers to the *Fourth set of replies* by Descartes (pp. 154–78).

Alanen, L., 1996. "Reconsidering Descartes's notion of the mind-body union." *Synthese* 106, 3–20.

Almog, J., 1986. "Naming without necessity." *The Journal of Philosophy* 83: 210–42.

Almog, J., 1991. "The what and the how." *The Journal of Philosophy* 88: 225–44.

Almog, J., 2002a. *What Am I? Descartes and the Mind-Body Problem*. New York: Oxford University Press.

Almog, J., 2002b. "Pains and brains," in S. Lindström and P. Sundström (eds.), *Physicalism, Consciousness, and Modality: Essays in the Philosophy of Mind*. Umeå Preprints in Philosophy series 2002: 1. Dept. of Philosophy and Linguistics, Umeå University.

Anderson, A. R., 1964. "Mathematics and the "language game"," in P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics*. Oxford: Basil Blackwell.

Baldwin, T., 2002. "Kantian modality." *Proceedings of the Aristotelian Society* supp. vol. 76: 1–24.

Balog, K., 1998. "Conceivability arguments." Ph.D. thesis, Rutgers University, New Brunswich, New Jersey.

Balog, K., 1999. "Conceivability, possibility, and the mind-body problem." *The Philosophical Review* 108: 497–528.

Barwise, J. and J. Etchemendy, 1992. *The Language of First-Order Logic*. CSLI Lecture Notes series 34. Stanford, CA: Center for the Study of Language and Information.

Bealer, G., 2002. "Modal epistemology and the rationalist renaissance," chapter 1 in T. S. Gendler and J. Hawthorne (eds.), *Conceivability and Possibility* (71–125). Oxford: Oxford University Press.

Belshaw, C., 2000. "Death, pain and time." *Philosophical Studies* 97: 317–41.

Bennett, J., 1994. "Descartes's theory of modality." *The Philosophical Review* 104: 639–67.

Bennett, J., 2001a. *Learning from Six Philosophers*, volume 1 (Descartes, Spinoza, Leibniz). Oxford: Clarendon Press.

Bennett, J., 2001b. *Learning from Six Philosophers*, volume 2 (Locke, Berkeley, Hume). Oxford: Clarendon Press.

Black, M., 1964. *A Companion to Wittgenstein's 'Tractatus'*. London: Cambridge University Press.

Block, N., 1981. "Introduction: What is the issue?" in N. Block (ed.), *Imagery* (1–18). Cambridge, MA: MIT Press.

Boghossian, P. A., 1996. "Analyticity reconsidered." http://www.nyu.edu/gsas/dept/philo/faculty/boghossian/papers/AnalyticityReconsidered.html (December 2002).

Casullo, A., 1979. "Reid and Mill on Hume's maxim of conceivability." *Analysis* 39: 212–19.

Casullo, A., 2000. "Modal epistemology: Fortune or virtue?" http://www.unl.edu/philosoph/ac_Spindel.htm (November 2001).

Chalmers, D. J., 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Philosophy of Mind series. New York: Oxford University Press.

Chalmers, D. J., 1999a. "Materialism and the metaphysics of modality." *Philosophy and Phenomenological Research* 59: 473–96.

Chalmers, D. J., 1999b. "Précis of *The Conscious Mind*." *Philosophy and Phenomenological Research* 59: 435–38.

Chalmers, D. J., 2002a. "Does conceivability entail possibility?," chapter 3 in T. S. Gendler and J. Hawthorne (eds.), *Conceivability and Possibility* (145–200). Oxford: Oxford University Press.

Chalmers, D. J., 2002b. "Consciousness and its place in nature." http://www.u.arizona.edu/~chalmers/papers/nature.html (October 2003).

Chisholm, R., 1979. "Identity through possible worlds: Some questions," chapter 3 in M. J. Loux (ed.), *The Possible and the Actual: Readings in the Metaphysics of Modality*. Ithaca: Cornell University Press.

Churchland, P. S., 1998. "The Hornswoggle problem," in J. Shear (ed.), *Explaining Consciousness: The Hard Problem*. Cambridge, MA: MIT Press.

Cottingham, J., R. Stoothoff, and D. Murdoch (eds.), 1984. *The Philosophical Writings of Descartes*, volume 2 (CSM 2). Cambridge: Cambridge University Press.

Cottingham, J., R. Stoothoff, D. Murdoch, and A. Kenny (eds.), 1991. *The Philosophical Writings of Descartes*, volume 3: The correspondence (CSMK). Cambridge: Cambridge University Press.

Dennett, D., 1981a. "The nature of images and the introspective trap," chapter 2 in N. Block (ed.), *Imagery*. Cambridge, MA: MIT Press.

Dennett, D., 1981b. "Two approaches to mental images," chapter 4 in N. Block (ed.), *Imagery*. Cambridge, MA: MIT Press.

Dennett, D., 1988. "Quining qualia," in N. Block, O. Flanagan, and G. Guzeldere (eds.), *The Nature of Consciousness*. Cambridge, MA: MIT Press.

De Pierris, G., 2002. "Causation as a philosophical relation in Hume." *Philosophy and Phenomenological Research* 64: 499–545.

Fine, K., 1994. "Essence and modality." *Philosophical Perspectives* 8: 1–16.

Fodor, J. A., 1981. "Imagistic representation," chapter 3 in N. Block (ed.) *Imagery*. Cambridge, MA: MIT Press.

Franzén, T., 1987. "Proofs and Provability." Ph.D. thesis, Dept. of Philosophy, Stockholm University.

Fricker, M., 1995. "Intuition and reason." *The Philosophical Quarterly* 45: 181–9.

Gendler, T. S. and J. Hawthorne, 2002. "Introduction: Conceivability and possibility," in T. S. Gendler and J. Hawthorne (eds.), *Conceivability and Possibility* (1–70). Oxford: Oxford University Press.

Gödel, K., 1964. "What is Cantor's continuum problem?" in P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics*. Oxford: Basil Blackwell.

Hale, B. 1996. "Absolute necessities." *Philosophical Perspectives* 10: 93–117.

Harman, G., 1967. "Quine on meaning and existence I." *Review of Metaphysics* 21: 124–51.

Hill, C. S., 1997. "Imaginability, conceivability, and the mind-body problem." *Philosophical Studies* 87: 61–85.

Hill, C. S. and B. P. McLaughlin, 1999. "There are fewer things in reality than are dreamt of in Chalmers's philosophy." *Philosophy and Phenomenological Research* 59: 445–54.

Hintikka, J., 1973. "Logic in philosophy—philosophy of logic," chapter 1 in *Logic, Language Games, and Information: Kantian Themes in the Philosophy of Logic*. Oxford: Oxford University Press.

Hintikka, J., 1998. "Epistemic logic." *Routledge Encyclopedia of Philosophy* CD-ROM, version 1.0.

Hintikka, J., 1999. "The emperor's new intuitions." *The Journal of Philosophy* 96: 127–47.

Hochberg, H., 1995. "Causality and generality in the *Treatise* and the *Tractatus*," chapter 80 in S. Tweyman (ed.), *David Hume: Critical Assessments*, volume 3. The Routledge Critical Assessments of Leading Philosophers series. London: Routledge.

Hume, D., 1975 [1777]. *Enquiries Concerning Human Understanding*, 3rd ed. by L. A. Selby-Bigge. Oxford: Oxford University Press.

Hume, D., 1978 [1739]. *A Treatise of Human Nature*, 2nd ed. by L. A. Selby-Bigge. Oxford: Oxford University Press.

Häggqvist, S., 1996. *Thought Experiments in Philosophy*. Ph.D. thesis, Dept. of Philosophy, Stockholm University. Stockholm: Almqvist & Wiksell.

Jackson, F., 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Clarendon press.

Kind, A., 1997. "The workings of the imagination." Ph.D. thesis, University of California, Los Angeles.

Kaplan, D., 1989. "Demonstratives," chapter 17 in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*. New York: Oxford University Press.

Kaplan, D., 1999. "Quantifying in," in A. P. Martinich (ed.), *The Philosophy of Language*, 2nd ed. Oxford: Oxford University Press.

Kneale, W., 1949. *Probability and Induction*. Oxford: Clarendon Press.

Kneebone, G. T., 1963. *Mathematical Logic and the Foundations of Mathematics: An Introductory Survey*. London: D. Van Nostrand Company Ltd.

Kripke, S., 1979. "A puzzle about belief," in A. Margalit (ed.), *Meaning and Use* (239–83). Dordrecht: Reidel.

Kripke, S., 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Körner, S., 1960. *The Philosophy of Mathematics*. London: Hutchinson & Co., Ltd.

Lewis, C. I., and C. H. Langford., 1959 [1932]. *Symbolic Logic*. Dover Books on Intermediate and Advanced Mathematics series. New York: Dover. Republication of 1932 ed.

Lewis, D., 1973. *Counterfactuals*. Library of Philosophy and Logic series. Oxford: Basil Blackwell.

Loar, B., 1999. "David Chalmers's *The Conscious Mind*." *Philosophy and Phenomenological Research* 59: 465–72.

Loux, M. J., 1998. *Metaphysics: A Contemporary Introduction*. Routledge Contemporary Introductions to Philosophy series. London: Routledge.

Mason, H. T. (ed.), 1985 [1967]. *The Leibniz-Arnauld Correspondence* (LAC). Philosophical Classics series. New York: Garland Publishing, Inc. Facisimile of Manchester: Manchester University Press, 1967.

Mason, R. V., 1988. "Logical possibility." *Metaphilosophy* 19: 11–24.

McGinn, C., 1976. "Anomalous monism and Kripke's Cartesian intuitions." *Analysis* 37: 78–80.

Melin, R., 1998. "Persons: Their identity and individuation." Ph.D. thesis, Umeå University. Umeå Studies in the Humanities series 143. Uppsala: Swedish Science Press.

Mellor, D. H., 1991 [1977]. "Natural kinds," chapter 7 in D. H. Mellor, *Matters of Metaphysics*. Cambridge: Cambridge University Press. Reprint from *British Journal for the Philosophy of Science* 28: 299–312.

Mill, J. S., 1974 [1843]. "A system of logic: Ratiocinative and inductive," chapter V in J. M. Robson (ed.), *Collected Works of John Stuart Mill*, volume 7, book 2, (224–51). London: Routledge & Kegan Paul.

Nagel, T., 1998. "Conceiving the impossible and the mind-body problem." *Philosophy* 73: 337–52.

Nagel, T., 2000. "The psychophysical nexus," in P. Boghossian and C. Peacocke (eds.), *New Essays on the A Priori* (433–71). Oxford: Clarendon Press.

Pap, A., 1958. *Semantics and Necessary Truth*. New Haven: Yale University Press.

Parsons, C., 1996. "Mathematical intuition," chapter 5 in W. D. Hart (ed.), *The Philosophy of Mathematics* (95–113). Oxford Readings in Philosophy series. New York: Oxford University Press.

Plantinga, A., 1974. *The Nature of Necessity*. Clarendon Library of Logic and Philosophy series. Oxford: Clarendon Press.

Plantinga, A., 1979. "Transworld identity or worldbound individuals?," chapter 7 in M. J. Loux (ed.), *The Possible and The Actual: Readings in the Metaphysics of Modality*. Ithaca: Cornell University Press.

Putnam, H., 1975. "The meaning of 'meaning'," in *Philosophical Papers*, volume 2: *Mind, Language and Reality*. Cambridge: Cambridge University Press.

Putnam, H., 1990. "Is water necessarily H_2O?," chapter 4 in *Realism with a Human Face* (55–79). Cambridge, MA: Harvard University Press.

Quine, W. V., 1947. "The problem of interpreting modal logic." *Journal of Symbolic Logic* 12: 43–8.

Quine, W. V., 1960. *Word and Object*. Cambridge, MA: MIT Press.

Quine, W. V., 1961a. "Reference and modality," in *From a Logical Point of View*, 2nd ed. (139–59). Cambridge, MA: Harvard University Press.

Quine, W. V., 1961b [1951]. "Two dogmas of empiricism," in *From a Logical Point of View*, 2nd ed. (20–46). Cambridge, MA: Harvard University Press.

Quine, W. V., 1990. *The Pursuit of Truth*. Cambridge, MA: Harvard University Press.

Quine, W. V., 1998. "Reply to Charles Parsons," in L. E. Hahn and P. A. Schilpp (eds.), *The Philosophy of W. V. Quine* (396–403). The Library of Living Philosophers series, volume 18. La Salle, IL: Open Court.

Rabinowicz, W., 2002. "Kripke on psychophysical identity," in S. Lindström and P. Sundström (eds.), *Physicalism, Consciousness, and Modality: Essays in*

the Philosophy of Mind. Umeå Preprints in Philosophy series 2002:1. Dept. of Philosophy and Linguistics, Umeå University.

Reid, T., 1994 [1863]. "Essays on the intellectual powers of man," in S. W. Hamilton (ed.), *The works of Thomas Reid*, volume 1, chapter B (215–498). Bristol: Thoemmes Press.

Ryle, G., 1984 [1949]. "Imagination," chapter 8 in *The Concept of Mind* (245–72). Chicago: University of Chicago Press. Reprint of London: Hutchinson, 1949.

Salmon, N. U., 1982. *Reference & Essence*. Oxford: Basil Blackwell.

Shoemaker, S., 1999. "On David Chalmers's *The Conscious Mind*." *Philosophy and Phenomenological Research* 59: 439–44.

Skyrms, B., 1989. "Tractarian nominalism," appendix in D. M. Armstrong, *A Combinatorial Theory*

Smith, B., 1992. "An essay on material necessity," in P. Hanson and B. Hunter, *Return of the A Priori*, Canadian Journal of Philosophy, supp. vol. 18 (301–22). Calgary: University of Calgary Press.

Soames, S., 1999. *Understanding Truth*. New York: Oxford University Press.

Sorensen, R. A., 1992. *Thought Experiments*. New York: Oxford University Press.

Sorensen, R., 2002. "The art of the impossible," chapter 9 in T. S. Gendler and J. Hawthorne (eds.), *Conceivability and Possibility* (337–68). Oxford: Oxford University Press.

Stalnaker, R. C., 1979. "Possible worlds," chapter 12 in M. J. Loux, *The Possible and the Actual: Readings in the Metaphysics of Modality*. Ithaca: Cornell University Press.

Stenius, E., 1996. *Wittgenstein's Tractatus: A Critical Exposition of the Main Lines of Thought*. Wittgenstein Studies series. Bristol: Thoemmes Press. Reprint of Oxford: Basil Blackwell, 1960.

Stoll, R. R., 1979. *Set Theory and Logic*. Dover Books on Advanced Mathematics series. Now York: Dover.

Stroud, B., 1995. "Hume and the idea of causal necessity," chapter 78 in S. Tweyman (ed.), *David Hume: Critical Assessments*, volume 3. The Routledge Critical Assessments of Leading Philosophers series. London: Routledge.

Tidman, P., 1994. "Conceivability as a test for possibility." *American Philosophical Quarterly* 31: 297–309.

Tye, M., 1991. *The Imagery Debate*. Representation and Mind series. Cambridge, MA: MIT Press.

Tye, M., 1995. *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Representation and Mind series. Cambridge, MA: MIT Press.

Van Cleve, J., 1983. "Conceivability and the Cartesian argument for dualism." *Pacific Philosophical Quarterly* 64: 35–45.

van Inwagen, P., 1998. "Modal epistemology." *Philosophical Studies* 92: 67–84.

Weigel, C., 2000. "On the relationship between conceivability and possibility." Ph.D. thesis, Temple University.

Wilson, M. D. 1978. *Descartes*. The Arguments of the Philosophers series. London: Routledge & Kegan Paul.

Wittgenstein, L., 1922. *Tractatus Logico-Philosophicus*. London: Kegan Paul, Trench, Trubner & Co.

Wittgenstein, L., 1958. *Philosophical Investigations*, 2nd ed. Oxford: Basil Blackwell.

Wittgenstein, L., 1964 [1956]. *Remarks on the Foundations of Mathematics*, ed. by G. H. von Wright, R. Rhees, and G. E. M. Anscombe. Oxford: Basil Blackwell. Reprint of 1956 ed.

Yablo, S., 1990. "The real distinction between mind and body." *Canadian Journal of Philosophy*, supp. vol. 16: 149–201.

Yablo, S., 1993. "Is conceivability a guide to possibility?" *Philosophy and Phenomenological Research* 53: 1–42.

Yablo, S., 1999. "Concepts and consciousness." *Philosophy and Phenomenological Research* 59: 455–63.

SUBJECT INDEX

Page number in italics indicate central explication of the corresponding term.

UMEÅ STUDIES IN PHILOSOPHY
Published by the department of Philosophy and Linguistics
Umeå University, Sweden

The series UMEÅ STUDIES IN PHILOSOPHY consists mainly in research communications (monographs, dissertations and collections of essays) in Philosophy from the Department of Philosophy and Linguistics at Umeå University. The series is not distributed through the bookstores, but individual volumes can be obtained from the department. Orders should be sent to Umeå Studies in Philosophy, Department of Philosophy and Linguistics, Umeå University, SE-901 87 Umeå, Sweden.

1. JONAS NILSSON: *Rationality in Inquiry*: *On the Revisability of Cognitive Standards*, 2000.
2. PER NILSSON: *Naturen, vetenskapen & förnuftet*: *Upplysningens dialektik och det andra moderna*, 2001.
3. JAYNE M. WATERWORTH: *Living in the Light of Hope*: *An Investigation into Agency and Meaning*, 2001.
4. ANDERS ODENSTEDT: *Cognition and Cultural Context*: *An Inquiry into Gadamer's Theory of Context-Dependence*, 2001.
5. RÖGNVALDUR INGTHORSSON: *Time, Rersistence, and Causality*: *Towards a Dynamic View of Temporal Reality*, 2002.
6. BENGT LILIEQUIST: *Ludwik Flecks jämförande kunskapsteori*, 2003.
7. PETER NILSSON: *Empathy and Emotions*: *On the Notion of Empathy as Emotional Sharing*, 2003.
8. ANDERS BERGLUND: *From Conceivability to Possibility*: *An Essay in Modal Epistemology*, 2005.