Mainstream and Formal Epistemology

VINCENT F. HENDRICKS
Roskilde University

Contents

1

Introduction

> It is a curiosity of the philosophical temperament, this passion for radical solutions. Do you feel a little twinge in your epistemology? Absolute skepticism is the thing to try . . . Apparently the rule is this: if aspirin doesn't work, try cutting of your head.
>
> Jerry Fodor (1985)

Humans are in pursuit of knowledge. It plays a significant role in deliberation, decision and action in all walks of everyday and scientific life. The systematic and detailed study of knowledge, its criteria of acquisition and its limits and modes of justification is known as epistemology.

Despite the admirable epistemic aim of acquiring knowledge, humans are cognitively accident-prone and make mistakes perceptually, inferentially, experimentally, theoretically or otherwise. Epistemology is the study of the possibility of knowledge and how prone we are to making mistakes. Error is the starting point of skepticism. Skepticism asks how knowledge is possible given the possibility of error. Skeptics have for centuries cited *prima facie* possibilities of error as the most substantial arguments against knowledge claims. From this perspective, epistemology may be viewed as a reply to skepticism and skeptical challenges. Skepticism is the bane of epistemology, but apparently also a blessing, according to Santayana (1955): "Skepticism is the chastity of the intellect, and it is shameful to surrender it too soon or to the first comer" (p. 50).

Skepticism is a tough challenge and requires strong countermeasures. In set theory, a powerful combinatorial technique for proving statements consistent with the axioms of set theory was invented by P. Cohen in the

1960s. The technique is called *forcing*. In particular, Cohen developed forcing in order to prove that the negation of the Axiom of Choice and the negation of the Continuum Hypothesis are consistent with the axioms of set theory. Today, there are various ways of using the forcing technique. One way is to construct an object with certain properties or to construct a model in which there are no objects with certain properties, thus forcing what you want directly – either constructing the object or iteratively destroying any such object.

Contemporary epistemologies have developed a family of countermeasures for standing up to the skeptical challenge; these exhibit a type of 'bluntness' similar to that of set-theoretical forcing.[1] The idea of epistemological forcing is as follows: whenever skeptics cite possibilities of error as arguments against knowledge claims, the strategy is to show that, although they are possibilities of error, they fail to be *relevant* possibilities of error. Some possibilities of error are simply not genuine – they are too remote, too speculative, or too much. These possibilities may accordingly be *forced* out and are henceforth not to be considered during the knowledge acquisition process. If the agent can succeed over the possibilities deemed relevant, then that is good enough for knowledge – knowledge will, or should, exhibit all the classical characteristics under forcing.

The influential *epistemic reliabilism* of Goldman, Nozick's elegant formulation of the *counterfactual epistemology* and Lewis's new *contextual epistemology* are all informal epistemological proposals observing the forcing relation.

Epistemic reliabilism (Goldman 1979, 1986) and especially the recent versions outlined in Goldman (1992, 1996) acknowledge the agent's limited cognitive abilities and accordingly deflate the agent's epistemic responsibilities. The idea is to replace the rather demanding requirements typically proposed by skepticism for justified knowledge possession with more lenient conditions. In principle, a particular justified belief may be false; however, its method or mode of acquisition must in general lead to true convictions. For knowledge to come about, besides the truth of the belief in question, its method of acquisition must rule out all relevant possibilities of error. The forcing technique is included in the method of acquisition. The method may not be able to exclude the possibility that Descartes' devious demon is feeding the agent systematically misleading sensations. Then again, this is not a relevant possibility of error. Or so

[1] Otherwise set-theoretical and epistemological forcing bear little resemblance to one another. In a certain sense one may even call them opposites. The term 'forcing epistemology' was coined in Hendricks 2001.

it is claimed. According to epistemic reliabilists, infallible methods are not required for knowledge. The development of epistemic reliabilism up to the current versions is scrutinized from the forcing perspective in Chapter 3.

Nozick's (1981) counterfactual reliabilistic knowledge definition, an adapted and supplemented version of a proposal put forth by Dretske (1970), is likewise a forcing proposal. The goal is to show that knowledge is in fact possible. The inherent decision procedure in Nozick's definition of knowledge, together with the counterfactual semantics, requires the agent to succeed in all possible worlds sufficiently close to the actual world. The agent may not know that he is not a brain in a vat – a famous thought experiment suggested by Putnam (1981) – but that possibility of error is so remote, and the semantics governing the counterfactual conditional guarantees the long distance. This counterfactual epistemology is the topic of Chapter 4.

Whereas both epistemic reliabilism and Nozickian counterfactual epistemology begin by confronting the skeptical challenge, Lewis's (1996) contextual epistemology, in contrast, assumes knowledge of a great many things in a variety of different contexts, particularly conversational contexts. 'Contextualists' hold the view that the standards for knowledge acquisition, possession and maintenance fluctuate with what is at issue – and at stake – in the particular linguistic context. The current interlocutors determine which possible worlds are real or relevant and also why and when. The knowledge that you are currently wearing sneakers may evaporate into thin air once you set foot in an epistemology class because in this new context you may doubt whether you even have feet to put your sneakers on. Be that as it may, we have knowledge, and epistemology starts from there – not from ignorance or demonstrations of the mere possibility of knowledge. Considering brains in vats and Cartesian demons is to 'epistemologize', which may make knowledge 'elusive' especially in an epistemology class. What is needed for obtaining knowledge are regulatives to rule out possible worlds dictated by the current (conversational) context and then describe how we avoid error and gain truth in the ones that are left. Contextual epistemology is discussed in Chapter 5.

It turns out that a host of *formal* epistemological proposals also share the forcing heuristics. Knowledge claims may be restricted by algebraic constraints defined for the accessibility relation between possible worlds, which is the forcing foundation for *epistemic logic* or *logical epistemology*. Logical epistemology originates with Von Wright (1951) and was propounded most notably by Hintikka (1962). The algebraic properties of the accessibility relation between possible worlds may sometimes be

defined in such a way that the skeptic has nowhere to go. The forcing characteristics and the (often neglected) epistemological significance of epistemic logic are the topics of Chapter 6.

Formal learning theory, also dubbed *computational epistemology* by Kelly (1996), focuses on the intrinsic solvability of inductive epistemic problems for both ideal and computationally bounded agents (Kelly 2000). The basic idea is that when an agent is faced with an epistemic learning problem, the problem determines a set of possible worlds in each of which the agent has to succeed to solve the problem and acquire knowledge. This is also forcing. Brains in vats sever the connection between knowledge acquisition and reliable inquiry, but short of that, agents may have quite a bit of reliable inductive knowledge. Although it is a logical paradigm, in that it utilizes tools from mathematical logic, it is also a procedural or effective paradigm, as it concentrates on learning and knowledge acquisition issues rather than modal operators, axiomatics and validity, as logical epistemology does. Computational epistemology is the topic of Chapter 7.

The last epistemological proposal to be considered is called *modal operator epistemology*.[2] Modal operator epistemology is a mixture of epistemic, tense and alethic logic and a few concepts drawn from computational epistemology. It was developed in order to study the validity of limiting convergent knowledge (Hendricks 2001). To obtain limiting convergent knowledge, the agent has to converge to the true hypothesis only in the possible worlds consistent with what has been observed so far. This approach also pays homage to the forcing relation. Brains in vats are as devastating here as elsewhere, but if blocked, knowledge may in the limiting end have a certain strength measurable by a yardstick devised by logical epistemology. An outline of the modal operator theory of knowledge, together with an analysis of its epistemological importance, is provided in Chapter 8.

Epistemology may be pursued in different ways:

- 'Mainstream' epistemology (which encompasses epistemic reliabilism, counterfactual epistemology and contextual epistemology) seeks necessary and sufficient conditions for the possession of knowledge using largely common-sense considerations and folksy examples and counterexamples (see Fig. 1.1).

[2] Elsewhere the paradigm is also known as *modal operator theory*, since the paradigm is flexible enough to study other modalitites than knowledge.

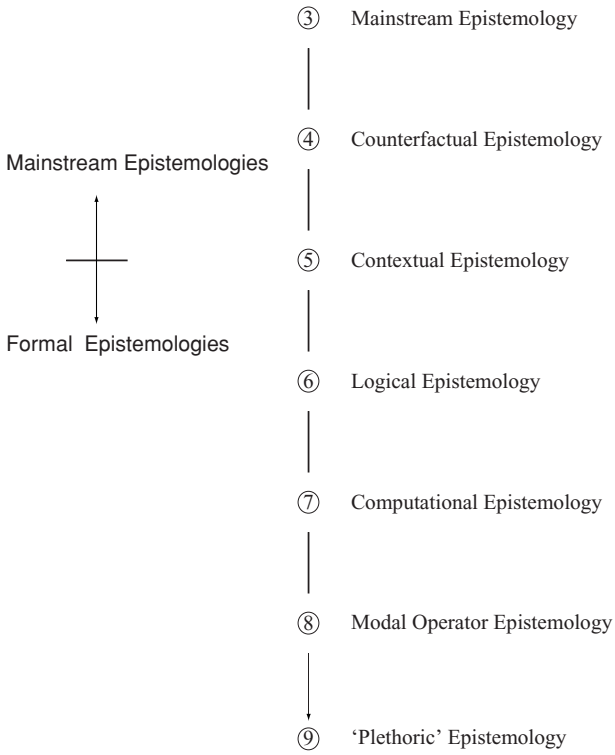

FIGURE 1.1. Epistemologies covered in this book and the chapters in which they are discussed.

- 'Formal' approaches to epistemology (which include logical epistemology, computational epistemology and modal operator epistemology) either proceed axiomatically or concentrate on learning and knowledge acquisition using toolboxes from logic and computability theory.

The two traditions have regrettably not paid much attention to each other. But both approaches, or rather their current exponents, employ the regulative forcing principle to combat skepticism. Based on this common denominator, the fundamental epistemological similarities and differences of the six paradigms may be cashed out in terms of how they each determine the set of possible worlds required for successful knowledge possession or acquisition.

The two approaches to the theory of knowledge share something else as well. One of the primary debates in contemporary epistemology

concerns the justification condition of the standard tripartite definition of knowledge as justified true belief. Time and time again, philosophers attempt to remedy the justification condition in order to avoid 'Gettierization' (Gettier 1963) and other epistemic unpleasantries. The justification condition is supposed to ensure that the belief and the truth conditions of the tripartite definition are 'adequately connected', that is, that the reasons for believing are truth-conducive and, insofar as they are, indicate what is meant by rational inquiry. Philosophy of science includes a subdiscipline concerned with exactly the same thing: methodology.

Methodology may crudely be characterized as the study of the methods by which science arrives at its posited truths. Methodologists and formally minded philosophers have a large technical toolbox available for analyzing and hopefully ensuring the truth-conduciveness of the methods of science. These techniques range from various inductive and nonmonotonic logics to Bayesianism, game theory and belief revision theory to formal learning theory, and so forth. When mainstream philosophers talk about justification, formalists speak of methodology. A philosopher may choose to invoke reliability; the formalist then asks how reliability is to be defined and what it can do for you methodologically. The mainstream epistemologist calls for a defeasibility condition, and the philosophical logician starts to think about default rules and nonmonotonic logics; the mainstreamer wants to get to the truth sooner or later, the computational epistemologist, say, begins to consider solvability and criteria of successful convergence; accumulating evidential support the mainstream community decides for and the Bayesian community will start conditionalizing; minimum mutilation of the web of beliefs and the belief revision theorists will work on revision functions and entrenchment relations; an epistemologist may worry about rationality, the game-theorist will start to consider, say, strategies for winning noncooperative games of perfect information. What the mainstream epistemologists are looking for may to some extent be what the formal epistemologists have to offer. But what the formal epistemologists have to offer the mainstream community, and vice versa being a two-way street, may also be quite sensitive to the perspectives on inquiry that the different approaches adopt.

The general prerequisites for studying these epistemo-methodological affinities are outlined in Chapter 2, then applied systematically in the subsequent chapters. Finally in Chapter 9, they are used for the purpose of outlining a program of 'plethoric' epistemology.

2

Priming the Pump

The epistemo-methodological prerequisites for comparing mainstream and formal epistemologies concentrate on the following items: the modality of knowledge, infallibility, forcing and the reply to skepticism; the interaction between epistemology and methodology; the strength and validity of knowledge; reliability; and the distinction between a first-person perspective and a third-person perspective on inquiry.

> If knowledge can create problems, it is not through ignorance we can solve them.
>
> Isaac Asimov

2.1 Modal Knowledge, Infallibility and Forcing

Agents inquire to replace ignorance with knowledge. Knowledge is a kind of epistemic commitment or attitude held toward propositions or hypotheses describing some aspect of the world under consideration.[1] Agents may in general hold a host of different propositional attitudes, such as belief, hope, wish, desire etc. But there is a special property that knowledge enjoys over and above the other commitments. As Plato pointed out, a distinct property of knowledge is truth. Whatever is known must be true; otherwise it is not knowledge, even though it very well may qualify as belief or some other propositional attitude.

Contemporary notions of knowledge are often *modal* in nature. Knowledge is defined with respect to other possible states of affairs besides the actual state of affairs (Fig. 2.1). The possibility of knowledge seems ruled

[1] The terms 'hypothesis' and 'proposition' are used interchangeably unless otherwise stated.

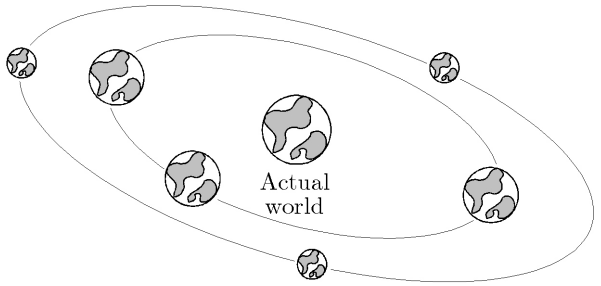

FIGURE 2.1. Modal knowledge is defined with respect to other possible worlds.

out when it is possible that we err. Introducing other possible state of affairs is an attempt to preclude exactly these error possibilities. Knowledge must be *infallible* by definition. As Lewis (1996) puts it, "To speak of fallible knowledge, of knowledge despite uneliminated possibilities of error, just *sounds* like a contradiction" (p. 367). A fallible notion of knowledge is not much different from a concept of belief potentially allowing the agent to 'know' a falsehood, severing the connection between knowledge and truth.

Plato also observed that knowledge, as opposed to mere true belief, is stable in nature. Knowledge has steadfastness and indefeasibility attached to it. True belief is quite useful as far as it goes, but in the light of true evidence, it may vanish. In the light of true evidence, knowledge will not evaporate. Inevaporability makes for the robust usefulness of knowledge compared with beliefs that are simply true. True belief in the actual world is not necessarily preserved if circumstances were to change but slightly. On the other hand, knowledge in the actual world is assumed to be stable across quite radically varying circumstances. Thus, among both informally and formally minded epistemologists, there is an agreement that knowledge is defined with respect to other 'possible worlds'. As Hintikka (2003a) notes,

> In order to speak of what a certain person *a* knows and does not know, we have to assume a class ('space') of possibilities. These possibilities will be called scenarios. Philosophers typically call them possible worlds. This usage is a symptom of intellectual megalomania. (p. 19)

There is an immediate difference between a philosophical logician and a philosopher. The logician typically remains agnostic about the ontological significance of the possible worlds and may refer to them as scenarios,

situations, states, contexts or conceptual constructions. The philosopher is usually quite concerned with the metaphysical baggage that comes along with the notion.[2]

Be that as it may, the stability and robustness of knowledge over other possible worlds leaves open the question of which space of worlds should be considered relevant for epistemic success. The classical conception of infallibilism is taken to require that an agent, in order to have knowledge of some hypothesis, must be able to eliminate *all* the possibilities of error associated with the hypothesis in question. The set of *all* worlds is considered. This set of possible worlds is too big for knowledge to have scope over. The set includes some rather bizarre worlds inhabited by odd beasts ranging from demons to mad and malicious scientists who have decided to stick your brain in a tank of nutritious fluids to systematically fool you. Or worlds in which contradictions can be true and tautologies can be false, like 'impossible, possible worlds' (Hintikka 1975). If these worlds were to be considered relevant all of the time, skepticism would have the upper hand all of the time. Epistemology may just end up with a fallibilistic notion of knowledge after all: there may not be a way for an agent to determine that he is not in the world of the beast or the brain. But then again, a fallibilistic notion of knowledge hardly qualifies as knowledge at all. At most, it amounts to a description of knowledge-seeking practices. Consequently, *if infallibilism is to be a viable reply to the skeptic, then infallibilism cannot be defined with respect to all possible worlds.* This is where epistemological forcing comes in.

The bizarre and extravagant possibilities of error may, under the right circumstances, be ignored even though they are logically possible, for all the agent knows. Knowledge may accordingly remain infallible but with world restrictions imposed. Forcing is more of an heuristic principle than an epistemological thesis proper:

> Whenever knowledge claims are challenged by alleged possibilities of error, the strategy is to show that the possibilities of error fail to be genuine in the relevant sense.

[2] To stay with currently adopted jargon, other scenarios, situations, states or contexts will be referred to as 'possible worlds' but nothing metaphysical is necessarily implied by the usage. Possible worlds are not to be understood as ontological or semantical totalities complete in their spatiotemporal history. Later it will become apparent that possible worlds may be endowed with enough formal structure to actually facilitate the achievement of important epistemological results.

Contemporary epistemologists choose to speak of the *relevant* possible worlds as a subset of the set of all possible worlds.[3] The philosophical logicians and other formal epistemologists consider an *accessibility* relation between worlds in a designated class within the entire universe of possible worlds. It will become apparent that there is no principled difference between relevance and accessibility. Informal epistemologies differ by the way in which relevance is forced given, say, perceptual equivalence conditions, counterfactual proximities or conversational contexts circumscribing the possible worlds. Formal epistemologies differ by the way in which the accessibility relation is defined over possible worlds. For example, philosophical logicians obtain different epistemic modal systems valid for a knowledge operator by varying (adding, dropping or relativizing) the properties of the accessibility relation, which might change from being reflexive and transitive to being reflexive, symmetric and transitive, for example.

Computational epistemology also forces as inductive epistemic problems to be checked for solvability specify a set of possible worlds for the agent to succeed over. Modal operator epistemology assumes that limiting convergent knowledge is restricted to possible worlds that are consistent with what has been so far observed.

> An informal epistemological paradigm may be a forcing strategy, and a formal one may be too. The task is then to find out how they each force so they can be put on a par for comparison.

Following Lewis (1996), one may say that it is a basic epistemic condition for agents to force for knowledge. The technically minded theorists of knowledge choose to formalize this basic epistemological condition.

2.2 Skepticism

The unstated premise of epistemology is, of course, that agents are seekers of knowledge or information – that is the whole point of conducting inquiry. Skepticism argues that even though gaining truth and avoiding error is the point of inquiry, the acquisition of knowledge is impossible given the standing possibility of error. In the end, we are left with ignorance rather than insight. Skeptics often cite two lines of argument in favor of their pessimistic conclusion.

[3] Explicit forcing proposals in the epistemological literature are sometimes referred to as 'relevant alternatives proposals'. Cf. Bernecker and Dretske 2000.

A great many skeptical arguments rely on the notion of *underdetermination*. In antiquity, skeptical arguments were molded around the possible fallibility of sense perception. The world is secretive in the sense that it does not always directly reveal what it is about. If the world does reveal what it really is like, this is not inconsistent with receiving scrambled signals. Any knowledge claim put forth by the agent about some aspect of the world is not guaranteed its truth if the truth depends on the 'underlying reality' or some other aspect transcending the immediate experience and evidence. Recent instances of such an argument may be identified in the well-known Cartesian demons, Hume's hidden springs of nature, the Duhem-Quine thesis, Kuhn's incommensurability, Putnam's brains in vats and Rorty's edification over inquiry.

The systematic underdetermination of a hypothesis by any available evidence is referred to as *global underdetermination* by Kelly (1996):

> A hypothesis is globally underdetermined if there are two possible worlds such that one of the worlds assigns the truth-value true to the hypothesis while the other assigns false in such a way that the evidence received by the agent remains the same forever independently of which of the two worlds is the actual world. (p. 17)

Global underdetermination leaves inquiring agents with two possibilities: one may either retreat to the Academic skepticism in which ignorance is bliss à la Carneades' and Arcesilaus's *ataraxia* (peace of mind) or invoke forcing to stay on the knowledge acquisition track. Forcing is obviously a way to deal with global underdetermination because global underdetermination only amounts to a skeptical objection if success is required in all possible worlds.

Skepticism plays on more than one string. Ever since the Pyrrhonian skepticism of Sextus Empiricus, the problem of induction and its various derivatives have presented a series of problems pertinent to knowledge possession and acquisition. The ornithologist may want to investigate whether all ravens are black. If he sets out to investigate by examining ravens one by one, there may not exist any finite specifiable time after the which the ornithologist can safely decide that all ravens are black. The next raven in line for observation could be white. By this argument, Sextus Empiricus was taken to have demonstrated the unreliability and consequently the untenability of inductive inference. Michel de Montaigne agreed with Sextus although Hume was awarded credit for the discovery, despite discussing causality rather than enumerative induction. A philosopher such as Popper identifies Hume as the one responsible for intellectual insight.

Skepticism about induction is the result of the *local underdetermination* possibly obtaining between evidence and hypothesis:

> A hypothesis is locally underdetermined by the evidence in a possible world if there is an infinite sequence of evidence possible for all the agent knows, such that each initial segment of this evidence sequence could arise independently of whether the hypothesis is true or false. (Kelly 1996, 24)

This definition implies the lack of a determinate finite time after which the agent can reach a decision concerning the truth or falsity of the hypothesis in question. An inductive skeptic, in turn, concludes that beliefs acquired by induction are locally underdetermined and cannot be justified.

Forcing is not going to come to the rescue either, or at least it should not. Eliminating the odd worlds does not preclude generalizations in those remaining unless one aspires to use forcing not only to dodge global underdetermination but also to boil down the set of relevant worlds in such a way that inductive inference essentially becomes deductive inference or something close to it (see Chapter 7).

There are sometimes cures for the problem of local underdetermination. If induction is unreliable, it may be the result of imposing a particularly demanding criterion of success for inductive inferences: *decision with certainty* demands that the agent, after having inquired for some period of time, will *halt*, conjecture the truth and do so in every possible world allowed by the forcing clause. This criterion of success looks like the one advocated by Plato in the *Theatetus*. Platonic inquiry has the agent stop after having inquired for some finite time. No matter what the truth will turn out to be, the agent will come to know what the truth is and will also know that he knows it. There is a finite time after which, for each later time, the agent has stabilized to the truth of some hypothesis.'Eureka!' the agent halts, outputs the truth and does not change his mind about it independently of the turn of the world. Not changing one's mind again during inquiry is referred to as *convergence*, and subsequently stopping inquiry as soon as convergence has arisen is called *convergence with certainty*.

Peirce (1958) thought that science in the long run may converge to the truth without ever producing an unequivocal sign of convergence. An agent may stabilize to the truth at some point but be unable to tell when stabilization occurred. There is convergence but no halting. Such a situation is referred to as *limiting convergence*. Given the evidence and forcing assumptions of the epistemic problem under consideration, computational epistemology has provided results to the effect that an inquiring agent may be logically reliable and guaranteed limiting arrival at the truth

for notions of successful assessment clearly weaker than decision with certainty (see Chapter 7). Epistemic problems of knowledge possession, assessment and acquisition may be reliably solvable for weaker notions of success and convergence. Beliefs acquired via inductive methods may sometimes be justified after all.

2.3 Epistemology

Both informal and formal epistemologies combat skepticism and the possibilities of error, but along different lines or paths. Since Plato's *Meno* and *Theatetus,* informal epistemology went down the path of identifying the defining ingredients of knowledge. Implicit in the works of Plato and in Kant's *Critique of Pure Reason* and explicitly described by C. I. Lewis (1946), knowledge is characterized by three individually necessary and jointly sufficient ingredients: truth, belief and justification. The *standard tripartite analysis of knowledge* still largely entertained in mainstream epistemology pays tribute to the idea that an agent Ξ, or an agent applying a method of inquiry, knows a hypothesis h insofar as the following conditions are satisfied:[4]

Ξ knows h iff
1. Ξ believes h,
2. h is true,
3. Ξ is justified in believing h.

The three distinct ingredients of the standard definition have not all enjoyed the same philosophical limelight. Believing is typically considered to be a psychological primitive or a dispositional psychological state. The state may exist independently of manifestation. Belief has not raised any serious eyebrows as an ingredient of knowledge. This first condition is essentially a condition hooking up the agent Ξ to hypothesis h. Knowledge presupposes belief that h, but meeting condition 1 alone is not sufficient for knowledge of h. The belief may turn out false. Humans are free to believe something which is in fact false, but knowledge is not such a lenient commitment. This paves the way for condition 2. The tripartite analysis accordingly suggests, as an additional necessary condition, that knowledge of h entails the truth of h. Truth has already been discussed as a necessary ingredient.[5]

[4] 'Agent' and 'method' are treated as one and the same unless otherwise stated.

[5] It should be observed that some philosophers, notably Williamson, have argued that the concept of knowledge is primitive. Knowledge is not composed of truth, belief and

To qualify as knowledge, not only must the belief be true but it must be held true in a *justified* or robust way. The justification condition has usually been in the limelight, from Plato to contemporary epistemology. It is not easy to define what is meant by justification. As Lenzen (1978) argues,

> Though there is basic agreement that *something* must be added to true belief to obtain knowledge, what precisely this 'something' is, remains far from being evident. Because of the vagueness of such notions as 'having sufficient reasons for believing', 'being justified in believing', it is difficult to make a decision concerning the adequacy of (5), *i.e.* that knowledge implies justification. (p. 28)

A claim to knowledge requires that the connection between the belief condition and the truth condition is 'adequate'. Yet the two conditions alone are jointly insufficient for the attainment of knowledge. True beliefs may be the result of blind luck, clairvoyance, random guessing, etc. Beliefs generated by such procedures do not amount to knowledge. The reason is that such procedures are rather obscure and seemingly unreliable, even if true beliefs have actually been produced. The adequate connection of condition 1 and 2 is severed, and condition 3 must therefore be instated. The condition is intended to provide supportive reasons explaining why the first two conditions are suitably connected. Only furnished with such supportive reasons, together with the satisfaction of the other two conditions, may the agent be said to have the necessary and sufficient testimony required for knowledge.

The term 'mainstream epistemology' refers to the *modus operandi* of seeking necessary and sufficient conditions for the possession of knowledge based on the standard definition or some close derivative thereof. The methodology involves advancing folksy and intuition-based examples and counterexamples, or if needed, less folksy and less intuitive examples and counterexamples. Reasons for the possession of knowledge may be undercut by 'suitable' counterexamples; these counterexamples may again be undercut by suitable counter-counterexamples, restoring knowledge, and so on. It is a dialectical and sometimes even 'diabolical' process that by its very nature balances between the theory of

justification but rather is a primitive (mental) state in the following sense: 'If the content of a mental state can depend on the external world, so can the attitude of that content. Knowledge is one such attitude. One's knowledge that it is raining depends on the weather; it does not follow that knowing that it is raining is not a mental state. The natural assumption is that sentences of the form "*S* knows *p*" attribute mental states just as sentences of the forms "*S* believes *p*" and "*S* desires *p*" do' (Williamson 2002, 6). I am indebted to Lars Bo Gundersen for pointing this out.

knowledge and skepticism. As odd as this way of conducting epistemology may seem, it should, to make sense, be viewed from the heights of a greater philosophical ambition: global conceptual understandings of various epistemic notions. These understandings are allegedly often gained by the delicate use of fictions, intuition pumps and thought experiments (Baggini and Fosl 2003).[6]

Some formal formats for epistemology may very well subscribe to the standard definition of knowledge. Others may not but may still have something epistemically pertinent to say. A logical epistemologist may agree that something must be added to belief to obtain knowledge. Whatever has to be added in terms of justification is, however, abundantly unclear, as Lenzen argued in the passage quoted earlier, so drop it because it is not pertinent to the aim of logical epistemology (other's like Hintikka [2003a] would disagree). Computational epistemology may be affiliated with a view of knowledge as reliably inferred stable true belief. It is forced neither to such a view nor to any other theses pertaining to the nature of knowledge. If knowledge, however, is about reliably extending truth with more truth, then computational epistemological investigations are of great relevance to such an epistemic conception, as Kelly (1996) noted. A final example: insofar as knowledge acquisition is construed as winning a game against nature or winning against a collective of other agents, game theory has something significant to offer pertaining to rationality as it relates to winning strategies. A winning strategy does not necessarily presuppose some particular understanding of knowledge, although a winning strategy may be sensitive to what knowledge (information) is available to the player about the other players (Osborne and Rubinstein 1994; van Benthem 2000).

Instead of pursuing a global conceptual understanding of knowledge, formal epistemologies proceed in a more piecemeal fashion. A certain amount of conceptual understanding is presupposed, or a certain set of conceptual parameters are fixed. The fixed parameters may be anything: the epistemic goal, the attitude, the method, the strength (validity) of knowledge/belief, the forcing relation, and so on, or some combination thereof. Such a fixation and the additional structure imposed by the

[6] Hintikka (1999a, 1999b) recently attacked the extensive use of 'intuitions' in contemporary philosophy. He tracks its recent history back, not to Gettier (see later), but nevertheless to the same period, when philosophers methodologically misunderstood Chomsky's transformational grammar program. Another source of misguided use of 'intuitions' around the same time was Kripke's determination of rigid designation, according to Hintikka.

formal apparatus of choice (logic, learning theory, probability theory, game theory, belief revision theory, etc.) give rise to a formal model of inquiry. In such a model, particular notions are naturally singled out for formal analysis, such as validity, reliability, computability, and rationality. The idea is then to study what follows from the model with respect to the concepts of interest. Although the results of the piecemeal conceptual analyses in a formal framework may not add up to a global concept of knowledge, they may all the same reveal something about the structure of the ingredients making up such a concept.

There may also be an initial operational difference between the mainstreamers and the formalists: whereas the former often remain quite vague about the tacit assumptions and presuppositions of their conclusions, which are based on intuitions and folksy examples, the latter use intuitions and examples only to illustrate results obtained within a framework. If intuitions are used for more than this, formal epistemologists are always required to state these usages explicitly as definitions, assumptions and lemmata. The explication is needed to evaluate the plausibility of the results subsequently obtained. Coming to terms with the operational discrepancy between mainstream and formal epistemologies is important for the realization of a plethoric epistemology. Neither the operational discrepancy nor the difference in local and global ambitions is an impediment to interactively glueing mainstream and formal epistemology together (see Chapter 9).

One may object that the endeavour of comparing mainstream and formal epistemologies based on a unilateral forcing heuristics is flawed from the outset. Let an operational difference between mainstream and formal approaches be granted. The real difference is that while formal epistemologies force by slamming the door shut on skepticism as a matter of initial assumption, mainstream proposals force by using 'possible but however unlikely' clauses.[7] This makes for a fundamental difference pertaining to how seriously the overall skeptical challenge should be taken and prevents the two epistemological approaches from being put on a par for comparison.

There are examples from formal epistemology to support a claim about the categorical dismissal of skepticism from the outset. Epistemic logicians sometimes assume no false beliefs and infallible knowledge for

[7] I am indebted to Charles Griswold, Juliet Floyd and Troy Catterson for bringing this objection to my attention during the discussion after my lecture at the Boston Colloquium (Boston University, February 21, 2003).

certain applications. There are other applications for which knowledge of knowledge is assumed but not infallible knowledge in terms of knowing one's ignorance. And there are yet other applications for which logical epistemology assumes knowledge to be even weaker, leaving the door open to skepticism more and more. It is dependent on the application, or rather the *context*, as to how strong knowledge is taken to be. This is, however, exactly Lewis's point in his mainstream contextualistic epistemology. Sometimes the door can be slammed shut on skepticism if the context is right. Nozick's mainstream counterfactual epistemology virtually subscribes to this when the demon world is too far off. Also to be taken into account is the fact that a formal approach like computational epistemology is to some extent skeptical in flavor, as underdetermination and the possibility of reliable inquiry are intimately connected (see Chapter 7). Mainstream and formal epistemologies seem to be on a par for comparative forcing analyses after all.

2.3.1 Interlude: Bayesian Epistemology

As opposed to the mainstream preoccupation with knowledge, some versions of Bayesian epistemology consider knowledge epistemically overrated and unnecessary for decision and action; measures of opinion are good enough as long as they are strong enough. Jeffrey (1992), for example, propounds a radical sort of Bayesianism that deflates the notion of knowledge and embraces only degrees of beliefs. Acceptance rules are rejected in part due to skeptical motives, and there is no certainty about anything save for logical laws (or rather, only probability 1 on these laws, and no certainty about anything at all). As radical probabilists reject certainty in any form, they shun knowledge as well. Less radical Bayesians make certainty and knowledge indistinguishable.

Bayesian epistemology may initially also be read as a formal forcing paradigm: it requires the agent to succeed in possible worlds with high prior probability and disregard sets of possible worlds with low probability (Earman 1992; Howson and Urbach 1989). If the truth is in a world ascribed 0 or infinitesimally small prior probability, the truth is never to be found. Bayesian epistemology does not alter the nature of the forcing relation. On this construal, Bayesian epistemology simply provides yet another argument for dumping error possibilities.

It is one thing to determine what knowledge is, its mode of justification and its potential resistance to skeptical challenges, it is another to study how knowledge or information may or should rationally change in light of what is already known and new evidence. Van Fraassen (1989) has

outlined the tenets of what he calls a 'new' epistemology, and he recommended a shift in epistemological focus from the 'defensive' *statics* of the classical definition of knowledge to the 'offensive' *dynamics* of epistemics and doxastics based on probabilities:

> What I hope for is some reconciliation of the diverse intuitions of Bayesians and traditionalists, within a rather liberal probabilism. The old we might call defensive epistemology, for it concentrates on justification, warrant for, and defense of one's beliefs.
>
> The whole burden of rationality has shifted from justification of our opinion to the rationality of change of opinion.
>
> This does not mean that we have a general opinion to the effect that what people find themselves believing at the outset is universally likely to be true. It means rather that rationality cannot require the impossible. We believe that our beliefs are true, and our opinions reliable. We would be irrational if we did not normally have this attitude toward our own opinion. As soon as we stop believing that a hitherto held belief is not true, we must renounce it – on pain of inconsistency! (p. 10)

This view, minus its inherent probabilism, is also shared by many contemporary pragmatists, from Levi to Putnam (in recent pragmatist incarnations) to authors influenced by cognitive science, like Gärdenfors. A common denominator for an instrumentalism like van Fraassen's and the pragmatism of, say, Levi is their joint emphasis on the idea that what is in need of justification is not static belief but the methods for changing it. This is what is common to various strands of this 'new' epistemology.

The central idea is to deny as irrelevant something that may seem presupposed by the very forcing metaphor, namely, that epistemology is largely conducted by engaging in a justificational game with the skeptic. The forcing metaphor may be taken to suggest that this is the essence of epistemology. Playing the game against the skeptic – which van Fraassen dismissively calls 'defensive epistemology' – is not to be understood as monopolizing all types of epistemological activity. Rather, one may have it both ways: 'forcing' and the 'new' epistemology are not mutually exclusive. Van Fraassen says that some of the new trends in epistemology focus on justifying the strategies for changing beliefs rather than on defending the beliefs that one happens to endorse at a certain instant. Reflecting on the methods for changing beliefs presupposes that one has some tools for representing belief and knowledge to begin with. Arguing for the appropriateness of this or that representation for solving one epistemic problem or the other may lead to interesting foundational issues, some of which might be connected with skepticism. Forcing may

sound defensive in this respect, but it comfortably leaves room for, endorses, and encourages the great variety of modern studies of dynamic epistemics and doxastics characteristic of a 'new' epistemology. In fact, by the end of the day, forcing may be viewed as a methodological feature rather than an epistemological one encompassing the statics as well as the dynamics of knowledge and other epistemic and doxastic attitudes relevant to epistemology (see Chapter 9).

Treating Bayesianism as another excuse for dismissing error possibilities is too simple a construal. It ignores a corpus of recent work based on nonstandard probabilities or on conditional probability. Arló-Costa (2001b) has argued that the most promising form of Bayesian epistemology is based on the idea of taking conditional probabilities as primitive. The resulting acceptance rules are capable of avoiding the lottery paradox (by implementing ideas inspired by the work of De Finetti and van Fraassen) and similar problems, and the conditional logic arising out of this form of unified probabilism seems to improve on other forms of radical probabilism.[8]

The literature on mathematical economics registers a great deal of work in this direction, regarding not only rules of acceptance but also rules of decision. Arló-Costa (2001b) recently made a distinction between 'monist' and 'pluralist' strategies in Bayesian epistemology. While the monist strategies only allow one probabilistic primitive (monadic or dyadic probability), the pluralist strategies accept doxastic primitives not reducible to probability (De Finetti, Levi, and others). These pluralist strategies in Bayesian epistemology are not immediately concerned with dumping error possibilities. They start with a notion of belief given as a primitive, which determines the space of probability carriers. This is a way of focusing on what Levi would call 'serious possibilities', but here the primary motivation is not 'defensive' but pragmatic. As stated by De Finetti (1974):

> In almost all circumstances, and at all times, we find ourselves in a state of uncertainty. Uncertainty in every sense. . . . It would therefore seem natural that the customary modes of thinking, reasoning and deciding should hinge explicitly and systematically on the factor uncertainty as the conceptually pre-eminent and determinative element. The opposite happens however: there is no lack of expressions referring to uncertainty, but it seems that these expressions, by and large, are no more than verbal padding. The solid, serious, effective and essential

[8] One can show that this form of unified probabilism is mappable to non-standard probability via the extension of a result first presented by McGee (1994) and Arló-Costa and Thomason (2001). I am indebted to H. Arló-Costa for pointing this out.

part of arguments, on the other hand, would be the nucleus that can be brought within the language of certainty – of what is certainly true or certainly false. It is in this ambit that our faculty of reasoning is exercised, habitually, intuitively and often unconsciously. (p. 24)

De Finetti makes clear that his set of certainties contains more than mere tautologies and that its main role is to determine or fix a space of possibilities:

Thinking of a subset of truths as given (knowing, for instance, that certain facts are true, certain quantities have given values, or values between certain limits, certain shapes, bodies or graphs of given phenomena enjoy certain properties, and so on), we will be able to ascertain which conclusions, among those of interests, will turn out to be – on the basis of the data – either certain (certainly true), or impossible (certainly false), or else possible. (p. 25)

When it comes to probability as such, De Finetti holds that 'probability is something that can be distributed over the field of possibility'.

Using a visual image, which at a later stage could be taken as an actual representation, we could say that the logic of certainty reveals to us a space in which the range of possibilities is seen in outline, whereas the logic of the probable will fill in this blank outline by considering a mass distributed upon it. (p. 26)

The initial certainties that reveal this space are not the propositions that can be defended against the skeptic. This body of certainties would be way too thin to engage in any kind of useful form of inquiry. They are the 'practical certainties' needed for everyday interaction and scientific inquiry. The agent considers himself as infallible at every instant about their truth. How to reconcile this with the need for changing views has been the main topic of an epistemology of authors like Levi.

2.3.2 Getting Gettier

A particularly devastating blow was directed at the established epistemological view in 1963. In a three-page paper, 'Is Knowledge Justified Belief?', Edmund Gettier gave the now legendary and quite scandalous counterexamples to knowledge as true justified belief. Russell (1956) had anticipated the counterexamples in the late 1940s:

It is clear that knowledge is a subclass of true beliefs. . . . There is a man who looks at a clock when it is not going, though he thinks that it is, and who happens to look at it at the moment when it is right; this man acquires a true belief as to the time of day, but cannot be said to have knowledge. There is the man who believes, truly, that the last name of the prime minister in 1906 began with a B, but who

believes this because he thinks that Balfour was prime minister then, whereas in fact it was Campbell Bannerman. (p. 170–1.)[9]

Even a stopped clock shows the right time twice a day. Looking at it just at one of those two moments does not suffice for knowing what time it is. Thus, Russell anticipated these problems, but the explicit formulation of the counterexamples are due to Gettier. The counterexamples have partly fixed the agenda for mainstream epistemological research since.

The Gettier paradoxes often involve the derivation of something true from something false. Smith may have in the past collected firm evidence that, together with other relevant background information, Γ, furnish supportive reasons for or perhaps even deductively entail the following hypothesis about Jones's possessions with respect to automotive vehicles:

h_1 : Jones owns a Ford car.

Suppose Smith has an acquaintance, Brown. Smith does not know where Brown is. For no particular reason Smith uses his internal randomizer and chooses a location, say, Boston, as Brown's current location. Then by applying the introduction rule for the disjunction, Smith concludes,

$$\frac{\Gamma \vdash \text{Jones owns a Ford car.}}{\Gamma \vdash \text{Jones owns a Ford car } \vee \text{ Brown is in Boston.}}$$

If the derivation is valid for Boston, it will be valid for Barcelona and Brest-Litovsk as well. Smith constructs the following three disjunctive hypotheses immediately entailed by h_1:

h_2 : Jones owns a Ford car $\vee$ Brown is in Boston.
h_3 : Jones owns a Ford car $\vee$ Brown is in Barcelona.
h_4 : Jones owns a Ford car $\vee$ Brown is in Brest-Litovsk.

Let Smith accept h_2, h_3 and h_4 based on his solid belief in h_1. Given the standard tripartite definition of knowledge, Smith is then justified in his belief in h_2, h_3 and h_4. He is cleared to consider them as instances of knowledge. Odd, since by randomizing Smith has no clue as to Brown's whereabouts, though that does not matter for the disjunctive truth. Gettier then tells this story: Jones in fact drives a rented car from AVIS so it is not his own Ford. By accident, however, Brown is in Barcelona. This information is still not available to Smith. It follows nevertheless that Smith is justified in believing h_3, but he does not know that h_3 is true. Smith has

[9] I am indebted to Robert van Rooij for bringing Russell's anticipation of the Gettier-examples to my attention.

gotten things right but for the wrong reasons: He claims to have knowledge because he apparently has reasons to believe that Jones owns a Ford, which he does not. All the same, truth is preserved because, unknown to Smith, Brown is in fact in Barcelona. Smith's original reasons for believing and being justified are undercut. According to the standard definition of knowledge, however, Smith may still be accredited with knowledge.

The other counterexample discussed by Gettier is based on two other celebrated logical principles – substitutivity of identicals *salva veritate* and the introduction rule for the existential quantifier. Suppose that Smith and Jones are applying for the same job. Suppose, once more, that given relevant background information and evidential support Smith is justified in believing

h_1 : Jones will get the job.

Smith is also justified in believing another hypothesis:

h_2 : Jones has \$10 in his pocket.

Given h_1 and h_2, the introduction rule for the existential quantifier and the principle of substitution of coreferential terms, Smith derives:

h_3 : The person who will get the job has \$10 in his pocket.

In the end, it turns out that Smith himself will get the job and, by the way, Smith himself has \$10 in his pocket. Smith is justified in believing the *de facto* true hypothesis h_3, but once again he is right wrongly.

One way to solve the Gettier paradoxes would simply be to reject the counterexamples and claim them defective on the basis that they rest on a questionable, perhaps even false, principle that false hypotheses can justify the agent's belief in other hypotheses. That is probably a bit too easy, and examples similar to the original ones contrived by Gettier have been produced that do not rely on this allegedly dubious principle.

The standard strategy has been either to put some more meat on the justification condition or supply a fourth condition to the standard definition to keep Gettierization from surfacing. A host of different fourth conditions have been proposed some of which have been able to solve the original Gettier problems, some of which have not, but either way they have usually occasioned yet new Gettier derivatives, some of them plausible, some of them less. In fact, the epistemological community has turned the production of Gettier derivatives into an industry; thus, only a few influential ones will be reviewed later. Finally, attention is also restricted to the additional conditions that in one way or the other attempt to solve the Gettier paradoxes by appealing to *reliability*.

The Gettier paradoxes are like a disease that is both virulent and primitive. They point to a fundamental flaw in the classical conception of knowledge and do so by using quite simple means, like the introduction rules for disjunction and the existential quantifier. A particularly malignant feature of the paradoxes is that they allow one to remain comfortably in the actual world. Given the truth conditions for the two disjuncts, one may extensionally compute the truth-value for the entire disjunction. One is not required to invoke intensional measures and other possible worlds for this computation. The same goes for the introduction rule for the existential quantifier and substitution *salva veritate*. In other words, the Gettier paradoxes do not require modal universes for their formulation, not even much tampering with the agent's local epistemic situation. Solutions to the Gettier paradoxes, however, most often rely on modal notions of knowledge. Though the paradoxes are virulent but primitive, the cures are often modal and complex, involving rather strong medicine. Both mainstream and formal approaches alike are licensed to prescribe such medicine.

2.3.3 From Justification to Methodology

Some mainstream epistemologists have attempted to account for the notion of justification in terms of *reliability* in order to solve the Gettier paradoxes. The idea is to block 'knowledge' for the wrong reasons by a reliable procedure that ensures knowledge for the right reasons:

- Goldman argues that the belief in some hypothesis is justified if and only if the method by which the belief is formed is reliable. Reliability means producing more true convictions than false ones in the actual world, sometimes in other worlds as well.
- Nozick insists on a 'heavier' strategy by appealing to a strongly reliable recursive procedure with a halting condition imposed in all nearby worlds.
- Lewis in his modal epistemology speaks of a rule of reliability to be enforced in all worlds deemed relevant by the current context. He refers in a footnote to a type of nomic sufficiency account of reliability à la Armstrong.

Regardless of what reliability is supposed to mean in the end, *it is a criterion imposed on the inquiry method or agent* responsible for producing the belief:

> The justification condition of epistemology ends up in methodology as the study of how science arrives at its posited truths, that is, how

beliefs are justified by the canons, norms or recommendations and internal workings of the method applied or the agent in question.

Justification may be vague and is in need of clarification. Clarification may come from consultations with methodology, as Sankey (1999) recently noted:

These are questions about the truth-conduciveness of the method. While they relate directly to the epistemic status of the method, they bear *indirectly on the nature of rational justification.* For if use of method conduces to truth, then, given the relation between method and justification, the warrant provided by the method is warrant with respect to truth. (p. 1)

Bonjour (1976), a mainstream epistemologist, expressed the same view:

An adequate epistemological theory must establish a connection between its account of justification and its account of truth: *i.e.* it must be shown that justification, as viewed by that theory, is truth-conducive, that one who seeks justified beliefs is at least likely to find true ones. (p. 75)

Methodological recommendations for 'rational' scientific inquiry, truth-conduciveness, reliability, convergence, strategies for winning games, changing beliefs economically and reliably, and the like, are at the very *core* of many formal epistemological proposals. Computational epistemology scrutinizes the feasibility of recommendations for getting to the truth reliably for both ideal and computationally limited agents; game theory models rationality among agents; belief revision theory concentrates on informational economy and the agent's rational change of beliefs; and so on.

An illustrative example of where formal epistemology meets mainstream epistemology occurs when Stalnaker (1996a) suggests using belief revision (and epistemic logic) to get out of the Gettier-cases [Stalnaker 1996a]: Milton knows that h iff Milton believes h and learning no further true information would lead him to change his mind. Given the standard analysis of belief revision, the analysis gives rise to an account of knowledge which validates the modal system **S4.2**.[10]

[10] I am indebted to Robert van Rooij for suggesting Stalnaker's solution as an example. Hintikka (1962) embraces a similar understanding of knowledge when he explains, 'If somebody says I know that p in the strong sense, he implicitly denies that any further information would have lead him to alter his view' (p. 21). Changing one's mind about p implies that one did not know p from the outset. This point about how knowing p implies that one would not change one's mind about p (which one may also find in some of Unger's skepticism-friendly early work) is what led Kripke to complain that if this were so, knowledge would demand doxastic intransigence – that is, if the agent knows p, then he must regard all information that suggests $\neg p$ as misleading. Kripke never

Mainstream epistemologists have usually been quite unimpressed by the epistemological results of the formal approaches. For instance, the results of modal validity for knowledge operators in epistemic logic have not been considered pertinent to the primary justificational concerns of mainstream epistemology. This is odd indeed, since epistemic and doxastic logics *are* the logics of knowledge and belief. Admittedly, some philosophical logicians have also been of the opinion that there is no, or should not be, any connection between epistemic logic and general epistemology and have accordingly blurred the connection even more.

Knowledge is to be justified, and by being justified, it may also attain a certain epistemic strength. Modal logic has devised a sliding scale of epistemic strength, since the modal systems that a knowledge operator may validate exactly says something about how strong knowledge is over other possible worlds. When justification may be accounted for in terms of belief revision, as Stalnaker suggests in the earlier example, knowledge attains a certain strength on the order of **S4.2,** which is demonstratively stronger than **S4** but weaker than, say, the modal system **S5**. Similarly in modal operator epistemology, if knowledge is defined as limiting convergence, then knowledge may be proved **S4** strong on the condition that the agent behaves in accordance with certain methodological recommendations (of justification). From the mainstream epistemological perspective, the formal validity of an epistemic operator is a yardstick for how strong a *modal* knowledge concept may become, how infallible it may be and how good a response epistemic strength is to skepticism. Some of the 'formal' mainstreamers, like Lewis and Nozick, actually discuss the modal systems, epistemic axiomatics and the modal strength of knowledge explicitly in their theories of knowledge. By way of example, Nozick views some of the closure conditions for knowledge, which are actually axioms of epistemic logic, as untenable because acceptance of these closure principles puts skepticism right back into play. Lewis then counters by claiming that knowledge is closed under implication and that one may all the same still leave the skeptic out of the game in a given uniform context.

2.4 Methodology

Methodology was described as the study of the methods and methodological recommendations by which science arrives at its posited truths.

actually committed his thoughts on this matter to print, like much of his thinking on epistemology, in fact, but others have cited them. I am indebted to Duncan H. Pritchard for directing my attention to this latter point.

A pertinent question in science and epistemology is *when* a method of inquiry is expected to have an answer ready on epistemic problems. In a previous section, two convergence criteria were introduced: (1) *convergence with certainty*, and (2) *convergence in the limit*. The first criterion requires the agent to succeed by some finite time and clearly signal this success by halting. The second criterion is weaker. It requires the agent to succeed by some finite time, but halting is not demanded.

A crisp formulation of certainty convergence may be found in Kelly (1996). Suppose Ξ is an arbitrary inquiry method that the agent applies, and let h be a hypothesis:

Ξ *converges to h with certainty iff there is a time n such that*

1. Ξ *signals at n that it is ready to conjecture,*
2. Ξ *conjectures h at* $n+1$*, and*
3. Ξ *does not signal earlier than n that it is ready to conjecture.*[11]

Convergence with certainty is viewed as the hallmark of convergence in epistemology and methodology. Due to Hume's problem of induction, hypothetico-deductivism is, for instance, committed to the bold formulation of universal hypotheses and to waiting for the incoming evidence to refute them. When a counterinstance is encountered, the hypothesis in question could not possibly be true. It is accordingly refuted with this type of certainty. An existential hypothesis has a similar property, but instead of being refutable, it is verifiable with certainty. Conjecture the existential hypothesis and wait for the first collaborating instance in the observed evidence. Eureka! The hypothesis is verified with certainty, so stop inquiry and output the truth.

As attractive as certainty convergence may be, it is not always possible to obtain this kind of security. Real epistemological and scientific problems are not always amenable to convergence with certainty. In these cases, one may choose to drop the stop condition but not the requirement of convergence. Limiting convergence emerges, as the agent is free to oscillate regarding his conjecture some finite number of times. This number is not specifiable in advance. At some point, nevertheless, the agent must reach a convergence modulus and stabilize his conjecture, even if he does not

[11] Note that immediately prior to the certain conjecture the method of inquiry is required to produce a *signal* (say, Eureka!) of certainty. This is due to the fact that the method may produce the sign of certainty more than once. Therefore, the certainty conjecture is taken to be the one following immediately after the first occurrence of Eureka!. Subsequent signals of certainty will be 'ignored, as though the method has finished its job and is merely producing irrelevant noise thereafter' (Kelly 1996, 48).

know, nor is required to say, when stabilization has occurred. American pragmatists like Peirce and James are sympathetic to this idea. As already briefly noted, Peirce held the view that it is impossible to say anything about the direction of science in the short run but that science may all the same asymptotically approach the truth in the long run. Similarly for James (1960), as knowledge of universal laws may become impossible to acquire if one is obligated to say when science has got it right.

Limiting convergence has become a more and more respected convergence criterion in philosophy; recent arguments for scientific realism in the philosophy of science rest on limit assumptions (Boyd 1984); computational epistemology utilizes limiting convergence for a variety of purposes, including the acquisition of certain characterization theorems for the inductive solvability of epistemic problems (Kelly 1996); and Bayesians apply a limiting convergence criterion to obtain 'almost sure' convergence theorems (Earman 1992).

Philosophy, of course, is a rather abstract field. Perhaps it is not too much of a surprise that limit considerations are to be found here. Computability theory and computational linguistics, say, are less abstract, but limit criteria nevertheless enter here as well. Gold (1965) observed that there is an interesting class of problems in computability theory that only can be solved:

> by infinitely long decision procedures in the following sense: An algorithm is given which, for any problem of the class, generates an infinitely long sequence of guesses. The problem will be said to be solved in the limit if, after some finite point in the sequence, all the guesses are correct and the same. (p. 28)

Limiting convergence may be defined in the following way:

> Ξ *converges to h in the limit iff there is a time n such that for each later time* n' Ξ *conjectures h at* n'.

An immediate question pops up: Why entertain a notion of convergence but no certainty of when convergence has occurred, as Kitcher (1993) asks:

> To be sure, there are [Bayesian] convergence theorems about the long run – but as writers from Keynes on have pointedly remarked, we want to achieve correct beliefs in the span of human lifetimes (p. 293).

The problem posed by local underdetermination is that one may not be sure to get it right in the 'span of human lifetimes'. To weaken convergence is not a way of cheating epistemology and science, for if the limit is what it in essence takes for truth, then rather wait around for

it. The short-run performance does not make us any better off. Peirce's asymptotic considerations pertaining to scientific progress are exactly intended to determine whether science is in progress. In the short run, no answer to this question can be had. Limiting convergence is actually also a condition of real scientific practice, as computational epistemologists like Martin and Osherson (1998) explain:

> The general point is that Ψ is not required to recognize or signal in any way that its conjectures have begun to converge. In this respect our paradigm is faithful to the situation of real scientists, whose theories remain open to revision by new, unexpected data. It is, of course, possible to define paradigms that require scientists to signal convergence. The prospects for success, however, are then diminished. (p. 12)

Using computational epistemological means, Schulte (2000), for instance, proves that the identification of conservation principles for particle reactions is a limiting tractable problem, not one tractable with certainty. To criticize a limiting solution for not being a solution with certainty is like criticizing an apple for not being an orange.

Whether a given epistemic problem is amenable to a solution depends not only on the convergence criteria used but also on what sort of success is envisioned. Convergence criteria by themselves say little about being right or wrong, correct or incorrect. Epistemic concepts like verification, refutation, decision and even discovery determine senses of cognitive success for methods of inquiry. Confirmation theorists have favored assessment methods of verification, falsificationists have preferred refuting assessment methods, and recent results of computational epistemology show how the discovery of new hypotheses from evidence may be successful.

If criteria of convergence are mixed with criteria of success, criteria of successful convergence are obtained, and these specify the senses in which epistemic problems could (or should) be solvable. Many mainstream epistemological paradigms seem to favor assessment in terms of decision with certainty. Goldman's epistemic reliabilism (Chapter 3), Nozick's counterfactual epistemology (Chapter 4) and Lewis's contextual epistemology (Chapter 5) all hold, for different reasons, that the agent's local epistemic environment suffices for deciding the truth status of the beliefs under scrutiny with certainty. The situation in logical epistemology is a bit trickier, since methods and successful convergence criteria are hardly discussed due to the initial inactive nature of the agents involved in the inquiry process. All the same, the philosophical motivations furnished for logical epistemology likewise indicate a tribute to certainty

convergence (Chapter 6). Computational epistemology concedes to the epistemological tradition that knowledge acquisition may indeed be possible for a variety of successful convergence criteria weaker than decision with certainty and also for discovery engines (Chapter 7). On these grounds, modal operator epistemology in turn utilizes limiting discovery as a success criterion for convergent knowledge (Chapter 8).

2.4.1 Methodology: Categorical or Hypothetical

James (1960) is responsible for the insight that there is a significant difference between avoiding error and gaining truth. One may avoid error while at the same time not necessarily gain truth. Some logical positivists thought they would gain the truth via their confirmational program without really checking. Worse, perhaps, it seems to be the case that many early proponents of confirmation theory, like Hempel, did not insist on convergence to a true hypothesis. Confirmation was end enough in itself. It follows, then, that one may, with or without a worry, confirm forever but at the same time head nowhere near the truth.

There exists a result due to Putnam (1963) but hardly noticed among philosophers of science that backs a worry. Putnam provided one of the first computational epistemological results: for any algorithm of extrapolation based on a Carnapian theory of confirmation, there is an epistemic problem that the Carnapian extrapolator cannot acquire knowledge of even when fed all possible instances of the problem. Using confirmation theory is not necessarily a truth-conducive methodological strategy. Computational epistemology has since then provided many striking results showing how norms of inductive rationality may interfere with reliability (see Chapter 7).

Epistemologists and methodologists are in the business of advocating guidelines for rational scientific inquiry in terms of truth-conduciveness. One has to make sure that the guidelines have the intended truth-conducive property, which is not always the case, as Swinburne (1968) demonstrates:

> Compatible with any finite set of phenomena there will always be an infinite number of possible laws, differing in respect of the predictions they make about unobserved phenomena. Between some of these ready experimental tests can be made, but experimental tests between others are less easy and between them we provisionally choose the simplest one. Evidence that a certain law is simpler than any other is not merely evidence that it is more convenient to hold that suggested law than any other, *but evidence that the suggested law is true.* (p. 51)

Swinburne's approach seems to exhibit an indifference to the question of whether proceeding in accordance with the prescription of simplicity

is truth-conducive. It is stipulated but not demonstrated. One may argue that simplicity is a guide to truth under special forcing stipulations such that all worlds are eliminated save minimally complex worlds consistent with the evidence. To ensure that simplicity is truth-conducive under these forcing clauses requires an analysis of reliability, but none seems forthcoming.

Arguing that principles of rational inquiry are normative in terms of getting to the truth should be done cautiously. The principles advocated may just be in the way of the truth, as shown by the existence of a method that violates the principles and could have gotten to the truth. One should accordingly distinguish between two types of methodology;

- In a categorical methodology, a methodological recommendation is advanced for its own end regardless of finding the truth.
- In a hypothetical methodology, a methodological recommendation is advanced with the aim of finding the truth.

2.5 Reliability

Reliability is a methodological property, not an epistemic goal as such. First, an inquiring agent may strike upon the truth in the actual world or relevant possible worlds without being reliable. Second, try replacing truth with reliability in the second condition of the standard tripartite definition of knowledge. It makes little sense to say that the hypothesis is reliable. It makes sense to say that the agent is reliably forming a hypothesis, but then it is a property an agent may enjoy and hence a justificational and accordingly a methodological concern.

One particularly important methodological concern is what the definition of reliability amounts to or should amount to. The community of epistemologists seems to be greatly divided on the issue. If knowledge is to be modal, infallible, and reliable to ensure truth-conduciveness, then reliability must be *categorical* in the following sense:

> A concept of reliability is **categorical** if defined in terms of unequivocal success for some convergence criterion in the actual world or relevant possible worlds.

Reliability is defined with respect to two parameters: sense of success and range of possible worlds. A familiar argument for categorical reliability is the following. If success is not unequivocal success, then one could end up in either one of the following situations: a situation in which

there are uneliminated possibilities of error or a situation in which a belief is 'reliably' formed but in fact false. The former is a situation in which reliability has scope over other possible worlds whereas the latter is a situation in which the forcing conditions may be strong enough to admit only the actual world. Unequivocal success implies an infallible notion of knowledge acceptable to both Nozick and Lewis, for instance.

On the other hand, a fallibilistic notion of knowledge is not unacceptable to Goldman, since he entertains a *stochastic* understanding of reliability:

> A concept of reliability is **stochastic** if defined in terms of a success to failure ratio for some convergence criterion in the actual or relevant possible worlds.

Goldman's primary reason for advocating a stochastic concept of reliability and possibly a fallibilistic notion of knowledge is simply that he considers categorical reliability to be too strong. It is an inadequate notion of reliability for actual epistemic practice. Nozick is perhaps less worried about describing actual epistemic practice and more concerned with prescribing medicine capable of curing recalcitrant diseases in epistemology, like the Gettier paradoxes (and other serious skeptical challenges). Lewis wants to have it both ways – practically executable prescriptions.

Categorical and stochastic reliability are linked to two sets of the distinctions: the distinction between descriptive and normative epistemology, and the distinction between a first-person and a third-person perspective on inquiry.

2.6 First- and Third-Person Perspectives

Contemporary epistemology often draws a distinction between descriptive and normative theories of knowledge. There is a similar distinction in moral philosophy between descriptive and normative ethics. The former attempts to describe actual moral behavior whereas the latter sets the standards for correct moral conduct.

Similarly, descriptive epistemologies account for actual epistemic practice whereas normative epistemologies prescribe rules of inquiry, including mechanisms for avoiding error and gaining truth, truth-conducive justification criteria, learning and winning strategies, procedures for revising beliefs, and so on. The distinction is sometimes blurred by the fact that when describing actual epistemic practice, one may have to define

various notions such as knowledge itself, justification and reliability, bringing normative aspects into the picture. Introducing a weak version of reliability like the stochastic one allegedly still rules out certain cognitive mechanisms – such as 'intuition', 'hunch' and 'clairvoyance' – as reliable means for getting to the truth. All the same, the agent in the environment may still cite them as reasons for holding convictions. Normative proposals may also exhibit descriptive characteristics if the means for solving some problem actually are available to the agent in his natural epistemic environment.

Both descriptive and normative epistemologies usually subscribe to the common premise that epistemic practice is in some sense 'rational'. What separates the two is their stance on whether epistemology is to simply describe epistemic practice or to try to optimize it. It would be irrational epistemic behavior to follow some practice demonstrated *a priori* to be en route to error (when this practice is available course of conduct to the agent in the environment). It is not necessarily irrational, on the other hand, not to follow some prescription if the natural epistemic milieu sets the standards for what the agent is able to do and this prescription is not among the available courses of action. The local epistemic circumstances may for one reason or the other bar the agent in question from choosing the best means for an end. The constraints could even be such that they reward 'irrational' behavior. Calling such situations irrational would undermine the common premise to which the two approaches subscribe. Not only may the environment limit the agent's behavior, other agents may as well. This is, for instance, illustrated by game theory's distinction between cooperative and noncooperative games.

Sometimes agents would be able to have more knowledge than they actually have if they were not constrained by their local epistemic milieu. Then they could freely pursue the optimal means for obtaining some desirable result, whether truth, epistemic strength or winning in some other sense. Thus, one may rightfully ask why epistemologists sometimes are in the business of means-ends prescriptions that no local agent is able to abide by. There are two related answers to this question:

- Epistemologists are in the business not only of ascribing knowledge to themselves but also of ascribing knowledge to other agents. Lewis (1996) has pointed out that there is a significant difference between one agent ascribing knowledge to himself in his local epistemic situation and others ascribing knowledge to him given the situation they are in. The two situations do not always coincide. Different persons do not share the same real world in many contexts. There are rules

to follow under knowledge attribution to oneself and others to know what we think we know.

- Rather a principled answer in the long run, than no answer in the short run; and rather, principled information about what it would take to solve the epistemic task at hand, than no information at all. Epistemology is about whether knowledge is possible and about what agents can and cannot know insofar as knowledge is possible. The problem is that it is not always clear from within whether something is knowable or not. One recurs to an outside perspective for a principled answer that may then spill over into the local circumstances.

According to Lewis, the agent may actually know more than we are able to ascribe to him. On his account, this is due to the fact that the attribution of knowledge is highly sensitive to which world is considered the actual to the agent in a given conversational context.[12] An agent in his environment is more likely to be aware of what the relevant possibilities are given the world considered actual by him than the knowledge ascriber standing by him or even outside. Lewis refers to these two stances as a *first-* versus a *third*-person perspective on inquiry.

Observe that an agent is free to be prescribe recommendations for himself to follow as long as the means suggested are available to him where he is. An outsider may also freely prescribe recommendations for the agent as long as they are available to the agent. If the outsider decides to prescribe a course of conduct to solve an epistemic problem for the agent in the environment but that course of conduct is unavailable to the agent, then the situation changes. What then becomes emphasized is what it would take to solve the epistemic problem regardless of whether the agent is capable of actually performing the necessary action(s).

The distinction between normative and descriptive and the distinction between first-person versus third-person perspectives are not mutually exclusive. The distinction between descriptive and normative theories of knowledge and a modified version of Lewis's first-person and third-person perspective dichotomy are subsumed in the following formulations:

> A perspective on scientific inquiry is **first-person perspective** if it considers what an agent can solve, can do or defend given the available means for an end and given the epistemic environment he is sunk into.

[12] What is also troublesome is that 'actual' is an indexical for Lewis (1984) in his modal ontology so there is not necessarily a real question of considering 'a world' actual. I am indebted to John Symons for reminding me of this feature.

Goldman's epistemic reliabilism and Nozick's counterfactual epistemology are thoroughly first-person perspectives, on this definition. Lewis's contextual epistemology encompasses both perspectives, as does logical epistemology, surprisingly enough, while computational epistemology and modal operator epistemology are more thoroughly third-person perspectives on inquiry.

> A perspective on scientific inquiry is a **third-person perspective** if it considers what an agent could solve, could do or defend given the best means for an end independently of the epistemic environment he is sunk into.

Just as with convergence criteria, one should be careful not to conflate the two different perspectives. Criticizing some position, whether mainstream or formal, without noticing that the criticism is based on a third-person perspective and that the position advocated is based on a first- person perspective may again turn out to be criticizing an apple for not being an orange. For example, distinguishing between the two perspectives can be of extreme importance for a formal epistemology like epistemic logic. The dichotomy has a significant bearing on the epistemology's general plausibility and its way of dealing with skepticism (see Chapter 6). Finally it will later become apparent that the distinction between the two perspectives is a *philosophical distinction that entails a formal difference.*

Distinguishing between these two perspectives on inquiry and observing, the other prerequisites highlighted in this chapter are the parameters with respect to which selected contemporary mainstream and formal theories of knowledge will be analyzed, assessed and compared. The parameters, at the same time, form parts of a program in 'plethoric' epistemology.