1

Rationality

INTRODUCTION

What is it for someone to be rational or reasonable, as opposed to being irrational or unreasonable? Think of some examples in which someone is being rational or reasonable as well as examples in which someone is being irrational or unreasonable. What do you think makes the difference? Think also of some examples in which someone makes a mistake but is not therefore irrational or unreasonable.

1.1.1 *Some Examples*

Here is one kind of example:

> *Giving in to temptation*
> Jane very much wants to do well in history. There is a crucial test tomorrow and she needs to study tonight if she is to do well in the test. Jane's friends are all going to a party for Bill tonight. Jane knows that if she goes to the party, she will really regret it. But she goes to the party anyway.

It is irrational for Jane to go to the party, even if it is understandable. The rational thing for her to do is to stay home and study.

Many examples of giving in to temptation involve a bit of irrationality. For example, smoking cigarettes while knowing of the health hazards involved is at least somewhat irrational. The rational thing to do is to give up smoking.

Here is a different sort of example:

> *Refusing to take a remedial course*
> Bob, a college freshman, takes a test designed to indicate whether students should take a useful remedial writing course. Students do not write their names in their examination booklets but write an identifying number instead, so that graders will not know the identity of

the students whose answers they are grading. Bob does poorly in the test and is told he should take a remedial writing course. He objects to this advice, attributing his poor score on the test to bias on the part of the grader against his ethnic group, and does not take the remedial writing course.

Bob's belief that his score is the result of bias is irrational. It would be more rational for Bob to conclude that he got a poor score because he did poorly on the test.

Refusing a reasonable proposal
Three students, Sally, Ellie, and Louise, have been assigned to a set of rooms consisting of a study room, a small single bedroom, and another small bedroom with a two-person bunk bed. Sally has arrived first and has moved into the single. The other two room-mates propose that they take turns living in the single, each getting the single for one-third of the school year. Sally refuses to consider this proposal and insists on keeping the single for herself the whole year.

Sally's room-mates say she is being unreasonable. (Is she?)

Confusing two philosophers
Frieda is having trouble in her introductory philosophy course. Because of a similarity in their names, she confuses the medieval philosopher Thomas Aquinas with the twentieth-century American philosopher W. V. Quine.

This is a mistake but does not necessarily exhibit irrationality or unreasonableness (although it may).

Failing to distinguish twins
Harry has trouble distinguishing the twins Connie and Laura. Sometimes he mistakes one for the other.

That by itself is not irrational or unreasonable, although it would be unreasonable for Harry to be over-confident in the judgement that he is talking to Connie, given his past mistakes.

Adding mistake
Sam makes an adding mistake when he tries to balance his chequebook.

A mistake in addition need not involve any irrationality or unreasonableness.

Consider mistakes about probability. Under certain conditions some people assign a higher probability to Linda's being a feminist and a bank teller than to her merely being a bank teller. The probabilities that people assign to certain situations can depend on how the situation is described, even though the descriptions are logically equivalent. Are mistakes of this sort always irrational or unreasonable? Are some of them more like mistakes in addition?

What is the difference between the sort of mistake involved in being irrational or unreasonable and other mistakes that do not involve being irrational or unreasonable? Does it matter what the difference is?

Do you think it is irrational or unreasonable to believe in astrology? To be superstitious? To believe in God? To believe in science? To be moral? To think that other people have mental experiences like your own? To suppose that the future will resemble the past? These questions increasingly raise a question of scepticism. A sceptic about X is someone who takes it to be irrational or unreasonable to believe in X. Is scepticism sometimes itself irrational or unreasonable?

1.1.2 *Rationality and Cognitive Science*

Issues about rationality have significance for cognitive science. For example, one strategy for dealing with cognition is to start with the assumption that people think and act rationally, and then investigate what can be explained on that basis. Classical economic theory seeks to explain market behaviour as the result of interactions among completely rational agents following their own interests. Similarly, psychologists sometimes explain 'person perception', the judgements that one makes about others, by taking these judgements to be the result of reasonable causal inferences from the way others behave in one's presence. In ordinary life, we often base predictions on the assumption that other people will act rationally (Dennett, 1971), as we do when we assume that other drivers will act rationally in traffic.

Such strategies require assumptions about rationality. Economics assumes that the rational agent maximizes expected utility (for example, von Neumann and Morgenstern, 1944). Classical attribution theory identifies rationality with the scientific method (for example, Kelley, 1967). It is less clear how we identify what is rational in our ordinary thinking. (One possibility is that each person asks what he or she would do in the other person's shoes and identifies that imagined response as the rational one.)

Some research has been interpreted as showing that people often depart systematically from the ideal economic agent, or the ideal scientist. People often ignore background frequencies, tend to look for confirming evidence rather than disconfirming evidence, take the conjunction of two claims to have a higher probability than one of the claims by itself, and so on.

There is more than one way to try to explain (away) these apparent departures from ideal rationality. One type of explanation points to resource limits.

> *Resource limits*
> Reasoning uses resources and there are limits to the available resources. Reasoners have limited attention spans, limited memories, and limited time. Ideal rationality is not always possible for limited beings: because of our limits, we may make use of strategies and heuristics, rules of thumb that work or seem to work most of the time, but not always. It is rational for us to use such rules, if we have nothing better that will give us reasonable answers in the light of our limited resources.

A second way to explain apparent departures from rationality is to challenge the view of rationality according to which these are departures even from ideal rationality. If people depart from what is rational according to a particular theory, that may be either because they are departing from rationality or because that particular theory of rationality is incorrect.

Some of the cases in which people appear to depart from ideal rationality are cases in which people appear to be inconsistent in what they accept. They make logical mistakes or violate principles of probability that they also seem to accept. How could these cases not be cases of irrationality?

Two ways have been suggested. First, it may be that people are not actually being inconsistent in their judgements.

> *Different concepts*
> People may be using concepts in a different way from the experimenter. When people judge that Linda is more likely to be a feminist bank teller than a bank teller, they may be using 'more likely' to mean something like 'more representative'. When people make apparent mistakes in logic, that may be because they mean by 'if' what the experimenter means by 'if and only if'. Given what they mean by their words, they may not be as inconsistent as they appear to be (Cohen, 1981).

Second, even if people are sometimes inconsistent, that does not show they are being irrational.

Reasonable inconsistency
It is not always irrational or unreasonable to be inconsistent (Pollock, 1991; Nozick, 1993).

It is an important question just what connection there is between being inconsistent and being unreasonable or irrational.

In this essay, I look more closely at rationality and reasonableness. I consider both actions and beliefs. What is it to act rationally or reasonably and what is it to act irrationally or unreasonably? What is it to have rational or reasonable beliefs and what is it to have irrational or unreasonable beliefs?

1.2 BACKGROUND

1.2.1 *Theoretical and Practical Rationality*

Let us begin by contrasting two of the examples mentioned above, 'Giving in to temptation' and 'Refusing to take a remedial course'. Jane goes to a party knowing she should instead study for tomorrow's exam. Bob thinks his grade on the writing placement exam is due to prejudice against his ethnic group even though he knows the grader does not have any way to discover the ethnic backgrounds of those taking the exam. One obvious difference is that Jane's irrationality is manifested in a decision to do something, namely, to go to the party, whereas Bob's irrationality is manifested in his belief, whether or not he acts on that belief. Bob does go on to make an irrational decision to refuse to take the writing course that he needs, but the source of that irrational decision is Bob's irrational belief. The source of Jane's irrational decision is not an irrational belief. Jane knows very well that she should stay home and study.

In deciding to go to the party knowing she should instead study for tomorrow's exam, Jane exhibits a defect in practical rationality. In believing that his grade on the writing placement exam is due to prejudice against his ethnic group, Bob exhibits a defect in theoretical rationality. Theoretical rationality is rationality in belief; practical rationality is rationality in action, or perhaps in plans and intentions.

Just as we can distinguish theoretical from practical rationality, we can distinguish theoretical reasoning, which most directly affects beliefs, from practical reasoning, which most directly affects plans and intentions. The upshot of theoretical reasoning is either a change in beliefs or no change, whereas the upshot of practical reasoning is either a change in plans and intentions or no change. Bob's irrationality arises from a problem with his

theoretical reasoning. There may be nothing wrong with his practical reasoning apart from that. Jane's irrationality arises entirely from a defect in practical reasoning and not at all from anything in her theoretical reasoning.

Theoretical and practical reasoning are similar in certain respects, but there are important differences. One important difference has to do with the rationality of arbitrary choices.

> *Arbitrary belief*
> Jane is trying to decide which route Albert took to work this morning. She knows that in the past Albert has taken Route A about half the time and Route B about half the time. Her other evidence does not support one of these conclusions over the other. So, Jane arbitrarily decides to believe that Albert took Route A.

Clearly, Jane should suspend judgement and neither believe that Albert took Route A nor believe that he took Route B. It is irrational or unreasonable for her to adopt one of these beliefs in the absence of further evidence distinguishing the two possibilities.

On the other hand, consider the practical analogue.

> *Arbitrary intention*
> Albert is trying to decide how to get to work this morning. He could take either Route A or Route B. Taking either of these routes will get him to work at about the same time and the balance of reasons does not favour going one way over going the other way. So, Albert arbitrarily forms the intention of taking Route A.

This arbitrary decision is quite reasonable. In fact, it would be quite irrational or unreasonable for Albert not to decide on one route rather than the other, even though his decision in the case must be completely arbitrary. Someone who was unable to make an arbitrary choice of routes would suffer from a serious defect in practical rationality! Arbitrary choices of what to intend can be practically rational in a way that arbitrary choices of what to believe are not theoretically rational.

Another difference between theoretical and practical rationality has to do with the rationality or irrationality of wishful thinking. Wishful thinking is theoretically unreasonable, but practically reasonable. Wishes and desires are relevant to practical reasoning in a way that they are not relevant to theoretical reasoning.

> *Wishful practical thinking*
> Jane's desire to get a good grade on the final exam leads her to study

for the exam in order to try to make it true that she will get a good grade on the final exam.

It is rational for Jane to let her desires influence her practical reasoning in this way. But consider the analogous theoretical case.

Wishful theoretical thinking
After Jane has taken the exam and before she has learned what her grade is, her desire to get a good grade on the exam leads her to conclude that she did get a good grade.

This sort of wishful thinking does not by itself give Jane a reason to believe that she got a good grade. To believe that something is so merely because she wants it to be so is theoretically unreasonable, whereas to decide to try to make something so because she wants it to be so is reasonable practical thinking. Desires can rationally influence the conclusions of practical reasoning in a way that they cannot rationally influence the conclusions of theoretical thinking.

This point has to be carefully formulated. Consider the following case in which desires do rationally influence what theoretical conclusions someone reaches.

Goal-directed theoretical reasoning
There are various conclusions that Jack could reach right now. He could try to figure out what Albert had for breakfast this morning. He could solve some arithmetical problems. He could work on today's crossword puzzle. He could try to resolve a philosophical paradox that Sam told him the other day. But, at the moment, Jack is locked out of his house and really ought to try to figure out where he left his keys. If Jack thinks about where he left his keys, however, he won't be able at the same time to resolve the philosophical paradox or solve the arithmetical puzzles. Because he wants very much to get into his house, he devotes his attention to figuring out where his keys must be.

Jack's goals can therefore be relevant to what conclusions he reaches. So, it is over-simple to say that your desires cannot rationally affect what conclusions you can legitimately reach in theoretical reasoning. Your desires can rationally affect your theoretical conclusions by affecting what questions you use theoretical reasoning to answer. The right statement of the constraint on theoretical wishful thinking therefore seems to be something like this: given what question you are using theoretical reasoning to answer, your desires cannot rationally affect what answer to that question you

reach. In practical reasoning, on the other hand, your desires can rationally influence not just the questions you consider but also the practical answers you give to those questions.

1.2.1.1 Practical Reasons for Belief

However, there are complications. Although wishful theoretical thinking is normally irrational, it is possible to have good practical reasons to believe something.

> *The power of positive thinking*
> Jonathan is sick. He has just read a study showing that people tend to recover more quickly if they believe that they will recover quickly. So Jonathan takes himself to have a practical reason to believe he will recover quickly.

> *Loyalty*
> Mary has been accused of stealing a book from the library. It would be disloyal for her best friend, Fran, to believe the charge against Mary. So Fran has a practical reason, loyalty, to believe that Mary is innocent.

> *Group think*
> Karen has been trying to decide what she thinks about capital punishment. She has noticed that the in-crowd at her school all believe that capital punishment for murder is justified and she has also noticed that members of the in-crowd do not like people who disagree with them about such things. Karen wants very much to be liked by members of the in-crowd. So she takes herself to have a practical reason to believe that capital punishment for murder is justified.

What do you think about this last example? Is there something wrong with Karen if she adapts her opinions to people she wants to please? How does that compare with Fran's belief in Mary's innocence based on loyalty to Mary?

Here are two further examples:

> *Advertising account*
> Landon would like very much to get the RST Tobacco advertising account. The RST Tobacco Company will hire only advertisers who believe that cigarette smoking is a healthy pastime. So Landon takes himself to have a practical reason to believe that cigarette smoking is a healthy pastime.

Pascal's argument for belief in God
Pascal (1995) reasons as follows. 'Either there is a God or there is not, and either I believe in God or I do not. So there are four possibilities with the following payoffs: (I) If I believe in God and there is a God, then I go to heaven and have infinite bliss. (II) If I believe in God and there is no God, then my costs are whatever is involved in believing in God. (III) If I do not believe in God and there is a God, then I go to hell and suffer the torments of the damned for eternity. (IV) If I do not believe in God and there is no God, then I have no costs and no gains. Now, the expected value of belief in God is the value of infinite bliss multiplied by the probability that there is a God minus the costs of belief in God multiplied by the probability that there is no God; and the expected value of not believing in God is the negative value of an eternity in hell multiplied by the probability that there is a God. No matter how small the likelihood that God exists, the expected value of belief is infinitely greater than the expected value of disbelief. Therefore, I should believe in God.'

Here we have what purport to be good practical reasons to believe one thing or another. This conclusion suggests that the difference between practical reasons and theoretical reasons is not just a matter of what they are reasons for, intentions *versus* beliefs.

1.2.1.2 Epistemic versus nonepistemic reasons for belief

All but the first of the examples in the preceding section have this feature: the examples mention a reason to believe something that does not make it more likely that the belief is true. Such reasons are sometimes called (for example, by Foley, 1987) 'nonepistemic reasons' for belief, in contrast with the more usual epistemic reasons for belief that do make the belief more likely to be true.

Epistemic reason for belief
R is an epistemic reason to believe P only if the probability of P given R is greater than the probability of P given not-R.

Nonepistemic reason for belief
R is a nonepistemic reason to believe P if R is a reason to believe P over and above the extent to which the probability of P given R is greater than the probability of P given not-R.

These definitions leave open the important question whether all practical

reasons for belief are nonepistemic reasons, a question we come back to below.

1.2.2 *Inference and Reasoning versus Implication and Consistency*

Issues about inference and reasoning need to be distinguished from issues about implication and consistency.

Inference and reasoning are psychological processes leading to possible changes in belief (theoretical reasoning) or possible changes in plans and intentions (practical reasoning). Implication is most directly a relation among propositions. Certain propositions imply another proposition when and only when, if the former propositions are true, so too is the latter proposition.

It is one thing to say

(1) A, B, C imply D.

It is quite another thing to say

(2) If you believe A, B, C, you should (or may) infer D.

Statement (1) is a remark about implication; (2) is a remark about inference. Statement (1) says nothing special about belief or any other psychological state (unless one of A, B, C has psychological content), nor does (1) say anything normative about what anyone 'should' or 'may' do (Goldman, 1986).

Statement (1) can be true without (2) being true.

Rationality versus genius
A, B, C imply D. Sam believes A, B, and C. But Sam does not realize that A, B, C imply D. In fact, it would take a genius to recognize that A, B, C imply D. And Sam, although a rational man, is far from a genius.

Here Sam has no reason at all to believe D. Consider also:

Discovering a contradiction
Sally believes A, B, C and has just come to recognize that A, B, C imply D. Unfortunately, she also believes for very good reasons that D is false. So she now has a reason to stop believing A, B, or C, rather than a reason to believe D.

Clutter avoidance
Jane believes A, B, C, she recognizes that A, B, C imply D, she does not

believe that D is false, and she has no reason to think that D is false. She is also completely uninterested in whether D is true or false and has no reason to be interested. D is the proposition that either $2 + 2 = 4$ or the moon is made of green cheese. There are many, many trivial consequences like this of her beliefs that she has no reason to infer. She has no reason to clutter her mind with trivial consequences of her beliefs just because they follow from things she believes.

Such examples indicate that, if implication is relevant to what it is reasonable to believe, the connection has to be fairly complex. (We discuss below how implication might be relevant to what it is reasonable to believe.)

Just as issues about implication have to be distinguished from issues about reasonable inference, issues about consistency have to be distinguished from issues about rationality and irrationality. Consistency and inconsistency are in the first instance relations among propositions and only indirectly relations among propositional attitudes. Propositions are consistent when and only when it is possible for them all to be true together. Propositions are inconsistent when and only when it is not possible for them all to be true together.

So, it is one thing to say,

(3) Propositions A, B, C are inconsistent with each other.

It is quite another to say,

(4) It is irrational (or unreasonable) to believe A, B, C.

The first remark, (3), unlike (4), says nothing special about belief or other psychological states, nor does it say anything normative. Hence, (3) can be true without (4) being true. Even if A, B, C are actually inconsistent, the inconsistency may have gone unnoticed and may be very difficult to discover. And even if you notice that A, B, C are inconsistent, there may still be reasons to accept each and it may be quite unclear which should be given up. You may not have the time or the ability to work out which should be given up or you may have more urgent matters to attend to before trying to figure out which to give up of A, B, C. In the meantime, it may very well be rational for you to continue to believe all three.

Age of the earth

In the nineteenth century, Kelvin's calculation of the age of the earth using principles of thermodynamics gave a result that was too small to allow for what was calculated to be the time needed for evolution (Gould, 1985). One scientific response was to continue to accept all

the relevant principles, despite their leading to this contradiction, while waiting for someone to figure out what was going wrong.

This would seem to have been a rational response to the difficulty. (Kelvin's calculations depended on assumptions about sources of energy. The discovery of radioactivity revealed a source he had not allowed for.)

Someone may show you a paradoxical argument leading to the conclusion that $3 = 1$, or a proof that a certain claim, which says of itself that it is not a true claim, is both a true claim and not a true claim.

Proof that 3 = 1.
Let $n = 1$.
Then $2n = 2$.
$n^2 + 2n = n^2 + 2$ [adding n^2 to both sides].
$n^2 = n^2 - 2n + 2$ [subtracting $2n$ from both sides].
$n^2 - 1 = n^2 - 2n + 1$ [subtracting 1 from both sides].
$(n + 1)(n - 1) = (n - 1)(n - 1)$ [factoring].
$n + 1 = n - 1$ [eliminating common factor from both sides].
$n + 2 = n$ [adding 1 to both sides].
$3 = 1$ [replacing n with its value, 1].

'Liar paradox'
Let (L) be the claim *that (L) is not true.*
The claim *that (L) is not true* is true if and only if (L) is not true [meaning of 'true'].
(L) is true if and only if (L) is not true [substituting].
But that is impossible [logic].

Someone can see that certain assumptions lead to paradox without being able to figure out which assumptions are most plausibly abandoned. In that situation, it may be rational to continue to accept the assumptions in question, trying to avoid the paradoxical patterns of argument.

1.2.3 *The Relevance of Goals and Interests*

The examples above called 'Goal-directed reasoning' and 'Clutter avoidance' indicate that what it is rational or reasonable for you to believe can depend upon your needs, goals, and interests in various ways. This is part of what lies behind the

General principle of clutter avoidance
It is not reasonable or rational to fill your mind with trivial conse-

quences of your beliefs, when you have better things to do with your time, as you often do.

If you spend all your time deriving trivial logical implications, for example, you will fail to attend to more important things, like finding food and drink and a place to spend the night.

More generally, whether it is rational to reach a particular conclusion will always depend in part on what questions you want to answer or have reasons to answer. If you need your keys to get into the house and you have data from which you could figure out where your keys are, then you have a reason to use those data to reach a conclusion about where your keys are. If it is urgent that you get into the house, it is not rational for you to spend your time drawing conclusions that do not promise to help you in this task. It is not rational for you to infer trivial consequences of your beliefs, as in '1 + 1 = 2; so either 1 + 1 = 2 or the moon is made of green cheese', even though the disjunctive proposition 'Either 1 + 1 = 2, or the moon is made of green cheese' has to be true if its first disjunct, '1 + 1 = 2' is true.

There is a practical aspect to all reasoning, including theoretical reasoning. What theoretical inferences it is reasonable for you to make depend in part on your needs and goals, because the inferences it is reasonable for you to make depend on what questions you have reasons to answer, and what those questions are depends on your needs and goals.

Of course, that is not to say that merely wanting P to be true can give you a reason to believe P (wishful theoretical thinking), although it may give you a reason to find out whether P is true, and it may give you a reason to make P true (wishful practical reasoning).

1.2.4 *Ideal Reasoners?*

Another point already mentioned is also behind the principle of clutter avoidance. Reasoning is subject to 'resource limits' of attention, memory, and time. So, it is not rational to fill your time inferring trivial consequences of your beliefs when you have more important things to attend to. Some theories of rationality (Stalnaker, 1984; Gärdenfors, 1988) begin by abstracting away from these limits. Theories of ideal rationality are concerned with an 'ideally rational agent' whose beliefs are always consistent and 'closed under logical implication'.

Deductive closure

An ideal agent's beliefs are deductively closed, or closed under logical

implication, if, and only if, any proposition logically implied by some of those beliefs is itself also believed.

Other theorists argue that such an idealization appears to confuse rationality, ideal or otherwise, with logical genius or even divinity! And, as we shall see, it is unclear how to relate such an ideal to actual finite human beings, with their resource-limited rationality.

We have already seen that ordinary rationality requires neither deductive closure nor consistency. It does not require deductive closure, because it is not always rational to believe D simply because D is implied by your beliefs in A, B, C. Rationality does not require consistency, because you can be rational even though there are undetected inconsistencies in your beliefs, and because it is not always rational to respond to the discovery of inconsistency by dropping everything else in favour of eliminating that inconsistency.

Now consider an ideal agent with no limitations on memory, attention span, or time, with instantaneous and cost-free computational abilities. It is not obvious whether such an agent would have a reason to infer all the trivial consequences of his or her beliefs. True, it would not cost anything for the agent to draw all those consequences, even all infinitely many of them, let us supposé. But there would also be no need to draw any of those consequences in the absence of a reason to be interested in them, for the agent can effortlessly compute any consequence whenever it may later be needed.

Could an ideal agent's beliefs be inconsistent? If these beliefs were also deductively closed, the agent would then believe everything, because everything follows from inconsistency.

Inconsistency implies everything
An inconsistent deductively closed agent believes both P and not-P.
Consider any arbitrary proposition Q.
P implies (P or Q), so the agent believes (P or Q).
Not-P and (P or Q) imply Q, so the agent believes Q.
So an inconsistent deductively closed agent believes every proposition Q.

Now consider rational recovery from inconsistent beliefs.

Ordinary recovery from inconsistency
An ordinary non-ideal rational agent, Tamara, believes that Bill is in his office, but when she looks into the office, no one is there. At least for a moment, Tamara has inconsistent beliefs, believing both that

> Bill is in his office and that no one is in Bill's office. Tamara quickly and painlessly recovers from this inconsistency by dropping her belief that Bill is in his office, concluding that he must have stepped out for a moment.

Ordinary rational agents deal with this sort of momentary inconsistency all the time, whenever something surprising happens. You are surprised when you believe P but discover Q, realizing that P and Q cannot both be true.

But consider the implications of surprise for an ideal deductively closed agent.

> *A deductively closed agent is unable to recover from inconsistency!*
> If the beliefs of such an agent were even momentarily inconsistent, the agent could never rationally recover, for there would be no trace in the agent's beliefs of how the agent had acquired inconsistent beliefs. Because rational recovery from inconsistency can appeal only to present beliefs, and, because the deductively closed agent has exactly the same beliefs no matter how he or she got into inconsistency, there is no way in which the deductively closed agent could use temporal criteria in retreating from inconsistency—the agent would have to recover in exactly the same way, no matter where he or she had started.

It is unclear how ideal rational agents might deal with ordinary surprise. Various possibilities suggest themselves, but we need not explore them here. In what follows, we will be directly concerned with real rather than ideal rational agents.

That is enough background. We now turn to some less obvious and more controversial aspects of rationality.

1.3 CONSERVATISM

The first less obvious aspect of rationality is that ordinary rationality is generally conservative in the following sense. You start from where you are, with your present beliefs and intentions. Rationality or reasonableness then consists in trying to make improvements in your view. Your initial beliefs and intentions have a privileged position in the sense that you begin with them rather than with nothing at all or with a special privileged part of those beliefs and intentions serving as data. So, for example, an ordinary

rational person continues to believe something that he or she starts out believing in the absence of a special reason to doubt it.

1.3.1 *Special Foundations: Rejection of General Conservatism*

An alternative conception of rationality going back at least to Descartes (1637) might be called 'special foundationalism'. In this view, your beliefs are to be associated with your reasons or justifications for them. These justifications appeal to other beliefs of yours, themselves to be associated with justifications, and so on. Circular justifications of belief are ruled out, so the process of justification ultimately rests on special foundational beliefs that are self-justifying and need no further justification. Special foundational beliefs include beliefs about immediate experience, such as headaches and perceptual experiences, obvious logical and mathematical axioms, and similar beliefs. In other words, you start from your evidence: those things that are evident to you. Rationality or reasonableness then consists in accepting only what can be justified from your evidence, on this view.

> *Ted's justification for believing that this is a piece of paper*
> It is thin, flexible, and white, with printing on it; it has the feel of paper rather than plastic. This evidence is best explained on the supposition that it is a piece of paper. Ted's justification for believing it is white is that it looks white to him and the circumstances of perception are such that something's looking white is best explained by the supposition that it is white. Ted needs no justification for believing that this looks white, because that is a foundational belief . . .

According to recent versions of special foundationalism (for example, Foley, 1987; Alston, 1989; Chisholm, 1982), foundational beliefs do not have to be guaranteed to be true. In the absence of specific challenges to them, they are justified, but their initial justified status might be overridden by special reasons to doubt them.

> *Defeating a foundational belief*
> Omar is terrified as he sits in the dentist's chair about to have a tooth drilled. When the dentist begins, Omar yells. The dentist stops and asks what's wrong. 'That hurt!' exclaims Omar, quite sincerely. 'But I haven't yet touched the drill to your teeth,' says the dentist. 'Oh!' says Omar after a pause, 'I guess I was confusing my anticipation of pain with actual pain.' Omar's initial foundational belief that he feels pain

is overridden by the further consideration that nothing had happened that could have caused pain. Beliefs about pain are foundational, but can be overridden by special reasons.

There are similar examples involving seemingly obvious logical or definitional truths.

Defeating a definitional belief
Paula is quite confident that all women are female, something she takes to be true by definition. Quinn objects, 'Wasn't a woman disqualified by the Olympic Committee for having the wrong chromosomes? Didn't they decide that she was not female?' Paula is set back by this question. 'I don't remember that case, but now that you mention that possibility, I can see that there could be a woman who is not, strictly speaking, female.'

Paula's confidence that she has intuited a definitional truth is shaken by the awareness of a possibility she had not previously considered. Seemingly obvious axioms or definitions are foundational but their justification can be overridden by special considerations.

We can describe each of the competing theories (foundationalism, conservatism) in the terminology of the other theory. So, we can say that the special foundations theory is conservative only about foundational beliefs. And we can say that general conservatism treats all beliefs as foundational.

1.3.2 *Objections to Special Foundationalism as a Theory of Rationality*

One problem for special foundationalism is to explain why special foundational beliefs should have special status. What distinguishes foundational beliefs from others that would justify applying conservatism to the foundational beliefs but not other beliefs?

A second, and perhaps more serious problem is that people tend not to keep track of their reasons for their nonfoundational beliefs. But, according to special foundationalism, if you don't associate a complete enough justification with a nonfoundational belief, then it is not rational or reasonable for you to continue to believe it. This realization may undermine a great many of your beliefs.

General beliefs with forgotten justifications
Foundationalist: What country is Athens in?
Maureen: That's easy—Greece. Everyone knows that!
F: But what reason do you have for thinking Athens is in Greece? Can

you remember a specific occasion on which you learned that information?
M: Well, no; but I'm sure if you just ask anyone . . .
F: But what grounds do you have now before you ask someone else?
M: I can't put my finger on anything specific, but I am sure.
F: If you don't have a justification that goes beyond the mere fact that you believe it, you are not justified in continuing to believe it.
M: Oh dear!

Specific beliefs originally based on perception
Foundationalist: Was Paul at the meeting yesterday?
Maureen: Yes, he was, although he didn't say anything.
F: Can you remember your perceptual evidence for thinking he was there?
M: Well, I remember seeing him.
F: Was he wearing a tie?
M: I don't recall.
F: Can you remember what he looked like?
M: Not in detail, but I do remember seeing him there.
F: If you no longer recall the sensory evidence on which that conclusion is based, you should abandon it.
M: That's ridiculous!

Originally, Maureen's belief was based on the evidence of her senses. But she almost immediately lost track of exactly what her sensory evidence was. Now she has at best the memory (another belief) that her belief was justified, without any special justification for it that would distinguish it from her other nonfoundational beliefs.

Special foundationalism implies that she should abandon such a belief as no longer justified. Because most of her nonfoundational beliefs are in the same position with respect to justification, almost all her nonfoundational beliefs should be abandoned as unjustified, according to special foundationalism. Special foundationalism implies that it is not reasonable or rational for her to continue to believe most of the things she currently believes! Some foundationalists are happy to endorse that sort of sceptical conclusion, but it is an extreme one and we will try to avoid such extremes in our discussion.

1.3.3 *The Burden of Proof*

The issue between general conservatism and special foundationalism amounts to a question about the burden of proof, or (better) the burden

of justification. According to special foundationalism, the burden of justification falls on continuing to believe something, at least for nonfoundational beliefs. Any nonfoundational belief requires special justification. Foundational beliefs do not require special justification. For them, what requires justification is failing to continue to believe them. Sometimes there is a reason to abandon a foundational belief, but such abandonment requires such a special reason.

According to general conservatism, the burden of justification is always on changing beliefs or intentions. You start with certain beliefs and intentions and any change in them requires some special reason. Any sort of change in belief or intention requires special justification. Merely continuing to believe what you believe or intend requires no special justification in the absence of a specific challenge to that belief or intention.

Which of these views, general conservatism or special foundationalism, best fits ordinary judgements about rationality and irrationality? (What do you think?) Not special foundationalism, for that view implies that it is irrational or unreasonable to continue to believe most of what you believe. So general conservatism fits better.

We now turn to a different issue, the relation between deduction and induction.

1.4 INDUCTION AND DEDUCTION

It is important to notice that deduction and induction are not two kinds of reasoning. In fact, induction and deduction are not two kinds of anything. Deduction is concerned with certain relations among propositions, especially relations of implication and consistency. Induction is not concerned with those or any similar sort of relation among propositions. Induction is a kind of reasoning. But, as we will see, deduction is not a kind of reasoning.

1.4.1 *Induction and Deduction as Two Kinds of Reasoning*

Consider this misleading account (based on Black, 1958*b*) of the relation between induction and deduction.

> *Deductive model of inference*
> Deductive logic is presented via a certain notion of 'proof' or 'argument'. A proof or argument has premises, intermediate steps, and a

> final conclusion. Each step must follow logically from prior steps in accordance with one or another specific rule, sometimes called a 'rule of inference'. Such a proof or argument is an instance of 'deductive reasoning'. Deductive reasoning in this sense is contrasted with 'inductive reasoning', which is said to take a similar form, with premises, maybe intermediate steps, and final conclusion, but with the following difference: deductive steps are always truth-preserving, whereas inductive steps are not.

This picture is very misleading. First, consider the reasoning that goes into the construction of a deductive proof or argument. Except in the simplest cases, the best strategy is not to expect to start with the premises, figure out the first intermediate step of the proof, then the second, and so on until the conclusion is reached. Often it is useful to start from the proposition to be proved and work backward. It is useful to consider what intermediate results might be useful.

The so-called deductive rules of inference are not rules that you follow in constructing the proof. They are rules that the proof must satisfy in order to be a proof.

In other words, there is a difference between reasoning that may involve the construction of a proof which must satisfy certain rules and reasoning that proceeds temporally in the same pattern as the proof in accordance with those rules. You do not reason deductively in the sense that your reasoning has the pattern of a proof. You can reason about a deductive proof, just as you can reason about anything else. But your reasoning is not well represented by anything like a proof or argument in the sense above.

1.4.2 *Implication and Consistency: Deduction*

Deduction is not a kind of inference or reasoning, although you can reason about deductions. Deduction is implication. A deduction or proof or argument exhibits an implication by showing intermediate steps.

Logic, the theory of deduction, is not by itself a theory of reasoning. In other words, it is not by itself a theory about what to believe (or intend); it is not a theory concerning how to change your view.

It is true that deductions, proofs, arguments do seem relevant to reasoning. It is not just that you sometimes reason about deductions in the way you reason about the weather or how much tax you owe. It is an interesting and nontrivial problem to say just how deductions are relevant to reasoning, a problem that is hidden by talk of deductive and inductive reasoning, as if it is obvious that some reasoning follows deductive principles.

The answer must be that it is often useful to construct deductions in reasoning about ordinary matters, and not just when you are explicitly reasoning about deductions or proofs. But why should it be useful to construct deductions? What role do they play in our reasoning?

Sometimes we do accept a conclusion because we have constructed a proof of it from other things we accept. But there are other cases in which we construct a proof of something we already accept in order to see what assumptions might account for it. In such a case, the conclusion that we accept might be a premise of the proof. The connection between proofs and reasoning is complex.

1.4.3 *Kinds of Induction*

The term 'induction' is sometimes restricted to 'enumerative induction'.

> *Enumerative induction*
> Given that all observed Fs are Gs, you infer that all Fs are Gs, or at least that the next F is a G.

But often the term 'induction' is used more widely so as to include also inference to the best explanation of the evidence.

> *Inference to the best explanation*
> Holmes infers the best explanation for the footprints, the absence of barking, the broken window: 'The butler wears size 10 shoes, is known to the dog, broke the window to make it look like a burglary . . .'

> *Scientific hypothetic induction*
> Scientists infer that Brownian motion is caused by the movement of invisible molecules.

What makes one hypothesis better than another for this purpose is something we must discuss later.

1.4.4 *Problem of Induction*

It is sometimes said that there is a 'problem of induction' (Bonjour, 1992).

> *(Alleged) problem of induction*
> When your beliefs logically imply the conclusion you come to accept, your conclusion cannot be wrong unless your premises are. Your premises guarantee your conclusion. This is not so in inductive

> reasoning, where your prior beliefs do not logically imply your conclusion. A question therefore arises whether you can be justified in drawing a conclusion that is not guaranteed by your premises.

But it is not clear what the problem of induction is supposed to be. Premises in an argument are to be distinguished from the starting points in reasoning, as we have already observed. The conclusion of an argument is not to be identified with the conclusion of reasoning in the sense of what you end up with or 'conclude' after reasoning. Even when reasoning culminates in the construction of an argument, the conclusion of the *argument* may be something you started off believing, and the conclusion of your *reasoning* may be to accept something that is a premise of an explanatory argument constructed as a result of inference to the best explanation.

Clearly, it would be stupid—indeed, highly irrational—not to engage in inductive reasoning. You would no longer be able to learn from experience. You would have no basis for any expectations at all about the future, for your evidence entirely concerns the past.

So, it would seem that the 'problem of induction' is a creation of confusion about induction and deduction, arising out of the deductive model of inference. Again, it is important to see that there are not two mutually exclusive kinds of reasoning, deductive and inductive. Deduction has to do with implication and consistency and is only indirectly relevant to what you should believe.

1.4.5 *Nonmonotonic Reasoning*

Unclarity about the relation between deduction and induction may be responsible for the occasional description of induction as 'nonmonotonic reasoning' in alleged contrast with deduction, which is described as 'monotonic'.

The terms 'monotonic' and 'nonmonotonic' are borrowed from mathematics.

> *Monotonic function*
> A monotonic (or 'monotonically nondecreasing') function $f(x)$ is a function whose value does not decrease as x increases. (A monotonic nonincreasing function is one whose value does not increase as x increases.) A nonmonotonic function is one whose value sometimes increases as x increases and sometimes decreases as x increases.

Deductive implication is monotonic in this sense:

Deductive implication is monotonic
Everything deductively implied by a set of propositions is also implied when additional propositions are added to a set. So, the deductive implications of a set of premises do not decrease in any respect as new premises are added. If A and B logically imply Z, so do A, B, and C, and so do A, B, C, and D, and so on.

On the other hand, reasoning is nonmonotonic in this sense:

Reasoning is nonmonotonic
Conclusions that are reasonable on the basis of specific information can become unreasonable if further information is added. Given the announced schedule for your course, your experience of the last few weeks, and that today is Monday, it may be reasonable for you to believe that your course will meet at 11:00 this morning. But if you are also given the further information that there is a sign on the classroom door saying that the 11:00 meeting of the course is cancelled today because your professor is ill, it is no longer reasonable for you to believe that your course will meet at 11:00 a.m. Now it is reasonable for you to believe that your course will not meet at 11:00 a.m. And, given the further information that the sign on the classroom door is a hoax by a student, it will be no longer reasonable to believe your course will not meet. New information can make old conclusions unreasonable, whereas additional premises in a deductive argument do not affect what conclusions follow deductively.

This aspect of inductive reasoning has been described in various ways. For example, it is sometimes said that inductive reasoning is 'defeasible'. Considerations that support a given conclusion can be defeated by additional information.

Sometimes this is described as 'default' reasoning. Given your original information, your default assumption is that the course will meet on Monday at 11:00 a.m. Additional information can override that default.

Default assumptions need not even be the usual case, as long as you can expect to find out when they do not hold. A default assumption might therefore take the form, 'Assume P, unless you hear otherwise.'

One use of default assumptions is sometimes called 'negation from failure'.

Negation from failure
The idea is to assume that something is not so unless you find information that it is so. Suppose, for example, you are interested in

whether there are any direct flights from Newark, New Jersey, to Lincoln, Nebraska. You do a computer search trying to locate such flights. When the computer does not find any, you conclude that there are none. The failure to find positive information leads you to accept a negative conclusion.

A number of attempts have been made to develop 'nonmonotonic logics' to capture these aspects of reasoning. Results have been fairly limited (Ginsberg, 1987). Some of these attempts are due to thinking of induction and deduction as two things of the same sort, the thought being that, because we have a deductive logic for deductive reasoning, we should develop an inductive logic for inductive reasoning. We have already seen what is wrong with this idea, namely, that deductive logic is concerned with deductive implication, not deductive reasoning. All reasoning is inductive.

It will be useful to develop an inductive or nonmonotonic logic only as an account of a kind of implication: default implication. Whether this development leads to any results that are useful to a theory of reasoning is still unclear.

There has been some discussion of the logic of conditionals, that is, statements of the form, 'If A, B'. At least some conditionals have the following sort of nonmonotonic property. 'If A, B' can be true, when 'If C and A, B' is not true.

Nonmonotonic conditionals
'If you turn the key, the engine will start' can be true even though 'If I disconnect the battery and you turn the key, the engine will start' is not true.

Horty and Thomason (1991) observe that research on the logic of conditionals comes together with research in nonmonotonic logic if we associate 'A default implies B' with 'If A, B'.

1.5 COHERENCE

The nonmonotonic aspect of inductive reasoning means that everything you believe is at least potentially relevant to the conclusions you can reasonably draw. Rationality is a matter of your overall view, including your beliefs and your intentions.

If it is reasonable to change your view in a certain way, let us say that your view would be more rationally 'coherent' if changed in that way. We

can then describe principles of rationality as principles of rational coherence.

Adopting this terminology, we can (following Pollock, 1979) distinguish two sorts of coherence, positive and negative.

1.5.1 *Negative Coherence*

Negative coherence is the absence of incoherence. Beliefs and intentions are incoherent to the extent that they clash with each other, for instance, through being inconsistent. Incoherence is something to be avoided, if possible, although we have seen that it is not always possible to avoid incoherence. Your beliefs might be inconsistent without your knowing that they are. And even if you are aware of inconsistency, you may not know of a sufficiently easy way to get rid of it.

> *Principle of negative coherence*
> To the extent that you are aware of incoherence in your view, you have a reason to modify it in order to get rid of the incoherence, if you can do so without too much expense.

Here is one way in which deductive logic is relevant to the theory of rationality, through providing an account of (one kind of) incoherence or inconsistency.

1.5.2 *Positive Coherence*

There is positive coherence among your beliefs (and intentions) to the extent that they are connected in ways that allow them to support each other. We can only speculate about what provides positive coherence. Some of the factors that seem relevant are the following.

> *Explanatory connections*
> A set of unrelated beliefs seems to be less coherent than a tightly organized conceptual scheme that contains explanatory and other principles that make sense out of most of your beliefs. This is why inference to the best explanation is an attractive pattern of inference.

Causal connections are a special case of coherence giving explanatory connections.

> *Causal connections*
> Belief in two events seems to be more coherent if one is seen as a cause

> of the other. When the lights go out in one room in her house, it makes more sense for Zelda to conclude that the fuse for that room has blown than to suppose that the fuse in a neighbour's house has blown. She easily envisages a causal connection between the fuse for that room blowing and the lights in the room going out. She does not as easily envisage a causal connection between the fuse in her neighbour's house blowing and the lights in her room going out.

To be sure, Zelda can envisage a complex causal connection between the fuse in her neighbour's house and the lights in her room. But to believe in that complicated connection would presumably offend against conservatism, which would seem to favour minimal changes in belief in order to obtain explanatory coherence. Also, without evidence of such complication, adding a belief in such a complication would actually decrease the overall coherence of her view.

Causation is not the only thing that would seem to bring explanatory coherence. Connecting generalization is another.

> *Coherence from connecting generalizations*
> All the emeralds Steve has observed are green. Steve infers that emeralds tend to be green, or even that all emeralds are green. This is an instance of enumerative induction.

We might think of enumerative induction as inference to the best explanation, taking the generalization to explain its instances. But then we must recognize that this is a different kind of explanation from causal explanation. A general correlation does not cause its instances!

Implication is an important kind of connector among beliefs.

> *Coherence from implication*
> Teri believes that Jack is either in his office or at home. She finds that his office is empty. She concludes that he is at home. This conclusion is implied by her prior beliefs.

Here is a second way in which deductive logic can be relevant to rationality. It is relevant to implication, and implication is a coherence-giving connection.

In trying to develop an account of rational coherence, we might try to reduce some of the factors mentioned to others in a substantive way. One idea would be to try to treat all factors as special cases of explanatory coherence. That idea is not very plausible for many cases like the last one, in which a conclusion is accepted because it is implied by other beliefs. What

is the relevant explanation in that case? One might say that the premises of Teri's argument explain why its conclusion is true. But that seems to stretch the notion of explanation.

Another idea would be to try to reduce all coherence to that involved in implication. That has some plausibility for certain explanations. And strict generalizations are related to their instances by implication. Often explanations in physics work via implication. Recognition of this fact gave rise to the so-called deductive nomological model of explanation (Hempel, 1965*a*), which works for many scientific explanations, but not for all.

One class of exceptions appeals to default principles that hold, other things being equal.

> *Explanation without implication*
> A certain substance dissolved in a certain liquid because it is sugar placed in water, and sugar normally dissolves in water. We have to say 'normally' because sugar does not always dissolve in water. It does not dissolve if there is already a supersaturated solution in the water, or if there is wax covering the outside of the sugar, or indefinitely many other things have occurred.

Here a general default principle helps to explain the dissolving in this case without guaranteeing that the sugar will dissolve. So, this explanatory connection is not based on strict implication.

1.6 SIMPLICITY

In trying to explain some data, it is reasonable to consider a very limited range of the infinitely many logically possible explanations. The rational inquirer restricts attention to the set of relatively simple hypotheses that might account for most of the data.

This is not to say very much, for it amounts to using the term 'simple' for whatever the relevant factors are that restrict rational attention to a certain few hypotheses. Furthermore, we are concerned with *relative* simplicity in this sense. A hypothesis that is too complicated, as compared with other available hypotheses at one time, can have a different status at another time if those other hypotheses have been eliminated. The first hypothesis might then be among the simplest of available hypotheses.

So, to say that the rational inquirer is concerned to find a simple hypothesis is not to say that the rational inquirer is committed to believing that 'reality is simple', whatever that might mean.

Let us now try to say more about simplicity in this sense, understanding that our discussion must be even more speculative than what has gone before. First, let us see how the relevant kind of simplicity might be involved in a famous philosophical 'riddle'.

1.6.1 *Goodman's 'New Riddle of Induction'*

Goodman (1965) discusses the following example. Suppose that Fran has a test for emeralds that does not depend on colour, she has examined various emeralds for colour, and she has found that each was green at least when she examined it. This evidence rationally supports the hypothesis

(H1) All emeralds are green.

Using the terminology of the preceding section, the evidence supports (H1) because it consists of instances of (H1) that are made more coherent if (H1) is true.

But there are many other hypotheses that are generalizations of the evidence, where the evidence consists of instances of each of these hypotheses. For example,

(H2) All emeralds are: either green if first examined before AD 2000 or blue if not first examined before AD 2000.

Goodman suitably defines the term 'grue' to stand for the predicate after the colon in (H2), so that the hypothesis can be abbreviated as follows:

(H2) All emeralds are grue.

Notice that (H2) conflicts with (H1) for any emeralds not first examined by AD 2000. According to (H1) those emeralds are green. According to (H2) they are blue.

Goodman points out that hypotheses like (H2) are not taken seriously. His 'new riddle of induction' asks what the difference is between (H1) and (H2).

Clearly, there is a sense in which Fran's (and our) preference for (H1) is due to its being a much simpler hypothesis than (H2). But what sort of simplicity is in question and why is it relevant?

1.6.2 *Using Simplicity to Decide among Hypotheses that are Taken Seriously*

It is very important to see that using simplicity to rule hypotheses out of consideration is to be distinguished from using simplicity as an explicit

consideration in theory choice. Sometimes a scientist will say that a particular theory is better than another because the first theory assumes the existence of fewer objects, fewer basic principles, or whatever. When a scientist argues in some such way he or she is arguing in favour of one rather than another hypothesis that is being taken seriously. As Sober (1988) observes, such appeals to simplicity are often quite controversial. That is, it is controversial whether simplicity in one or another respect is a relevant consideration in choosing among hypotheses.

But, even where there are deep controversies in a subject, reasonable disputants will still take seriously only a very few of the infinitely many possible hypotheses. We are concerned with whatever it is that leads reasonable people to disregard most of the hypotheses as too 'silly' to be considered.

(To repeat an earlier point, silliness is a relative matter. Hypothesis (H2) is silly because (H1) has not been ruled out. We can imagine a situation in which (H2) becomes acceptable.)

Let's call the sort of simplicity we are concerned with 'basic simplicity'. Because the phenomenon of ruling out crazy or silly hypotheses occurs in all domains, let us assume that there is a single domain-independent notion of simplicity for this purpose.

1.6.3 *Speculation: Basic Simplicity Has to Do with How Easy it is to Use Hypotheses*

The basic simplicity of a hypothesis seems to have something to do with the simplicity of its representation. But it is always possible to represent any hypothesis simply, so the matter is a bit more complex.

> *Simple representation of a complex hypothesis*
> We have already seen that the complex hypothesis (H2), 'All emeralds are: either green if first examined before AD 2000 or blue if not first examined before AD 2000', can be given a much simpler representation, if a suitable predicate is defined: 'all emeralds are grue.'

In fact, any hypothesis can be abbreviated by a single symbol, so simplicity of representation cannot be taken at face value.

Now, if a hypothesis like 'All emeralds are grue' is used to explain the data, it has to be expanded to its more complex form, 'All emeralds are: either green if first examined before AD 2000 or blue if not first examined before AD 2000.' This expansion is required on the assumption that we are more interested in accounting for the colours of objects, like whether they are blue or green, as opposed to their 'cholers', like whether they are grue

or bleen. If instead we were more interested in explaining why emeralds were grue, we could use the hypothesis 'All emeralds are grue' without having to expand it, and the hypothesis 'All emeralds are green' would require elaboration in terms of 'grue' and 'bleen' in order to provide the desired explanation.

So, perhaps the thing to look at is not so much the mere statement of the hypothesis but also how complicated it is to use the hypothesis to explain the data and predict new observations of a sort in which we are interested. (Here again theoretical rationality would depend on practical concerns.)

> *Simplicity as ease of use*
> In considering possible explanations of given data, it is rational and reasonable to ignore hypotheses that are much harder to use in explanation and prediction than other available hypotheses that in other respects account equally well for the data.

1.6.4 *Parasitic Theories*

A parasitic theory says that, as far as evidence goes, it is as if some other theory were true.

> *Descartes's demon hypothesis*
> Your sensory experience is the result of a powerful evil demon, giving you experiences as if of a world of physical objects.

> *Scientific instrumentalism*
> Scientific theories can be used as devices for calculating observations, but should not be treated as saying anything about the real nature of the world. All that can be rationally believed is that it is as if this or that scientific theory holds (van Fraassen, 1989).

In the classroom, it may be unclear how you can reject Descartes's demon hypothesis. But it would be crazy to take that hypothesis seriously in ordinary life. Similarly, outside the philosophy classroom it makes sense to take scientific instrumentalism seriously only when a theory can be accepted as no more than an instrument; for example, when the theory is known not to be wholly true. In that case, it makes sense to consider instrumentalist hypotheses.

> *Newton's laws as instruments*
> Relativity theory tells us that Newton's laws are not completely accurate, but they hold as good approximations at speeds much less than

the speed of light. Under those conditions, it is as if Newton's laws were correct.

We do not take parasitic theories seriously unless we have reason to reject the theories on which they are parasitic. In other words, parasitic theories are treated as 'less simple' than the theories on which they are parasitic.

This result fits our tentative suggestion that simplicity should be measured by how easy it is to use a hypothesis to explain data and make new predictions. A parasitic theory is normally more complicated according to this suggestion than is the theory on which it is parasitic, because to use the parasitic theory you have to do everything you do when using the nonparasitic theory and you have to do something more. You first calculate what is to be expected on theory T, then use the principle that what will happen is what is expected according to theory T. So, there is an additional step to the use of the parasitic theory that is not part of the original theory T.

Nonparasitic explanation
Why does E occur? Because of initial conditions C and laws L. Given C and L and the following calculation . . ., we expect E.

Parasitic explanation
Why does E occur? According to theory T, it is because of initial conditions C, and laws L. Given C and L and the following calculation ..., we would on theory T expect E. Our theory is that things will occur as if T is true. So, we expect E also.

The explanation of E from the nonparasitic explanation occurs as a part of the parasitic explanation. So, the parasitic explanation has to be somewhat more complicated than the nonparasitic explanation.

1.7 PRACTICAL RATIONALITY AND REASONABLENESS

So far, all that has been said about practical rationality is that your goals play a role in practical rationality that they do not play in theoretical rationality. The negative part of this remark, concerning theoretical rationality, may require qualification, given the apparent role of simplicity and conservatism in theoretical rationality, if these factors have a practical justification. We will discuss the possible need for such a qualification in the next section. In the present section, we say something more about the way in which goals are relevant to practical rationality.

One issue is whether there is a single category of goal, or perhaps a single measure of 'utility', as opposed to a variety of functionally different things: desires, values, goals, intentions, commitments, principles, rules, and so on. A related issue is whether we need to allow for a structure within goals in which some goals depend on others.

But let us begin with a few remarks about the mathematical decision theory that is often used as a model of rationality in economics.

1.7.1 *Decision Theory*

In its simplest form (for example, von Neumann and Morgenstern, 1944), mathematical decision theory applies when you are faced with a decision between two or more exclusive acts. Each act has one or more possible outcomes to which you assign certain values or 'utilities'. Let us use $u(A)$ to represent the utility of act A. You also assign conditional probabilities, $p(O,A)$, to each possible outcome O in relation to a given act A. Then the 'expected gain' of a given outcome O of an act A is $u(O) \times p(O,A)$. The 'expected utility' of each act A is the sum of the expected gains of each possible consequence of that act. Finally, the theory holds that rationality requires doing either the act with the highest expected utility or, if there is a tie for highest, one of the acts with highest expected utility.

The principles of decision theory are like principles of logic in being principles of consistency or coherence. It would be a mistake to identify decision theory with a full theory of practical rationality, just as it is a mistake to identify the theory of theoretical rationality with logic.

Some decision theorists argue that it is useful for individuals faced with hard practical problems to think of them in decision-theoretic terms. Such individuals are advised to consider carefully what their possible acts are, what possible consequences each act might have, what utility they assign to each possible consequence, and how likely they think a given act would be to have a given consequence. They should then calculate expected utilities and choose that act with the highest calculated expected utility.

Is that good advice? That is an empirical question: do people do better using such a method or not? The suggested method is not obviously good advice. Given a poor enough assignment of utilities and probabilities, you could be led very wrong by your calculation.

1.7.2 *Derivative Goals*

Some goals are derivative from others in a way that is important for practical rationality. You want A. B is a means to A. So you want B. That is, you

want B as a means to A. If you get A in some other way, you no longer have the same reason to want B. Or if you discover that B is not going to lead to A you no longer have the same reason to want B. It is irrational to continue to pursue an instrumental goal after your reason for wanting it has lapsed.

Also, consider the problem of deciding what to do when you have several goals. If you do A, you will satisfy goals G1, G2, and G3. If you do B, you will satisfy goals G4, G5, and G6. It is not easy to say how a rational person reaches an overall evaluation of acts A and B by combining his or her evaluation of the outcomes of each act. One idea (Franklin, 1817) is to try to reduce the lists by trying to match outcomes of A with equivalent outcomes of B, cancelling these equivalent goals out, and then considering only the remaining advantages of each course of action. That can still leave difficult choices.

But one thing can be said: do not count the satisfaction of two goals as distinct advantages of an act if your only reason for one of the goals is that it will enable you to attain the other.

> *Choosing a career*
> Mabel is trying to decide between a career in business and a career in teaching. These careers are associated with different lifestyles, and she considers which lifestyle she would prefer. She also considers the difference in income and wealth associated with the two choices, forgetting that income and wealth are means to the lifestyles associated with the choices.

Mabel is irrationally counting the same consideration (style of life) twice when she treats income as a separate consideration.

1.7.3 *Nonultimate, Noninstrumental Desires*

You can care about things that are neither ultimate ends nor instrumental toward getting other things you want.

> *Good news*
> Jack has been tested to see whether he has a fatal disease D. The test is quite reliable. Jack desperately wants the results of the test to be negative, indicating that he does not have the disease. Jack's desire is not an ultimate end of his, nor is it a desire for something that might be instrumental in obtaining something else that Jack desires. He desires a negative result because of what it indicates about him, not because of what it might lead to.

Notice that Jack's desire in this case is not for something that he could rationally treat as a goal. It would be irrational for Jack to bribe a lab technician to guarantee that the test yields a negative result. That wouldn't have any effect on whether Jack has disease D, which is (after all) what Jack is basically concerned with.

1.7.4 *Intentions*

Does a rational person always reason directly from current goals, always figuring out the best ways to maximize satisfaction of current goals? That would resemble special foundationalism with respect to theoretical reasoning.

It ignores the role of long-term intentions. Such intentions record the decisions already made. These decisions are not irrevocable, but they carry considerable weight and should not be frivolously discarded. A person incapable of maintaining long-term intentions would be incapable of long-term planning and would have at best only a low level of rationality (Bratman, 1987).

Intentions are not reducible to desires and beliefs, but put constraints on current planning of a special kind. A person's actual goals, as contrasted with things merely valued or desired, might be identified with what that person intends.

Intentions are directly related to action in ways not fully understood. Some authors think there are special intentions to do something now, constituting acts of will or volitions serving as the immediate causes of action.

1.7.5 *Strength of Will*

Our initial example of irrationality was an example of practical irrationality: Jane goes to the party rather than study for her exam. She finds the immediate pleasure of an evening more attractive than the longer-term considerations involved in doing well in her history course.

It is not that Jane temporarily overvalues the immediate pleasure of the party and undervalues the longer-term gains of study. She remains aware of the relative importance of these things. Her desires conflict with her evaluations.

In such a case, rationality requires sticking with her previously formed intentions, staying with her principles and resisting temptation.

1.7.6 *Reasonable Cooperation*

Finally, consider our earlier example of unreasonable negotiation, which I repeat:

> *Refusing a reasonable proposal*
> Three students, Sally, Ellie, and Louise, have been assigned to a set of rooms consisting of a study room, a small single bedroom, and another small bedroom with a two-person bunk bed. They discuss the proposal that they should take turns, each getting the single for one-third of the school year. Sally refuses to consider this proposal and insists on keeping the single for herself the whole year.

When her room-mates say that Sally is being unreasonable, they seem to be making a moral judgement about Sally. She is not being 'fair' (R. W. Miller, 1992).

Notice that her room-mates say that Sally is being 'unreasonable' and would not say that she is being 'irrational'. Similarly, a teenager asking for permission to use the family car might plead with his mother by saying, 'Be reasonable, Mom!' and not by saying, 'Be rational, Mom!'

1.8 THEORETICAL RATIONALITY AND PHILOSOPHICAL PRAGMATISM

Earlier I said that goals are relevant to practical rationality in a way in which they are not relevant to theoretical rationality. Although your goals are relevant to what questions it is rational for you to be interested in answering, they are not relevant to determining the answer you should accept through theoretical reasoning in the way in which your goals can be relevant to determining what it is rational for you to decide to do through practical reasoning. Wishful thinking is theoretically irrational even as it is practically okay.

We mentioned the possibility of good practical reasons to believe certain things and were therefore led to distinguish epistemic or theoretical reasons to believe something from nonepistemic practical reasons to believe something. Evidence that John was elsewhere at the time of the crime is an epistemic or theoretical reason to believe him innocent. On the other hand, loyalty to John provides a nonepistemic, practical reason to believe him innocent.

The possibility of philosophical pragmatism complicates this picture.

Everyone can agree that practical considerations are relevant to the choice of a notation for developing a theory.

> *Roman numerals*
> It would be hard to balance your bank account if you had to use roman numerals rather than the more standard arabic decimal notation. There are good practical reasons to use the one notation rather than the other.

Philosophical pragmatism argues against any sharp distinction between choice of theoretical hypothesis and choice of notation (Quine, 1960*a*). Pragmatists stress such practical features as we have already mentioned—simplicity, ease of use, and conservatism, for example—in deciding what to believe about any subject.

But then what happens to the distinction between theoretical and practical reasoning or, more precisely, the distinction between epistemic and nonepistemic reasons?

Pragmatists can still allow for this last distinction, defined as we defined it earlier.

> *Epistemic reason for belief*
> R is an epistemic reason to believe P only if the probability of P given R is greater than the probability of P given not-R.

> *Nonepistemic reason for belief*
> R is a nonepistemic reason to believe P if R is a reason to believe P over and above the extent to which the probability of P given R is greater than the probability of P given not-R.

Considerations of simplicity and conservatism are reflected in our probability judgements in a way that more specific practical considerations are not. For example, of the hypotheses that explain the evidence, we treat the simpler hypotheses as more likely to be true than the less simple hypotheses, given that evidence. On the other hand, a rational advertising agent should not suppose that it would be evidence that cigarettes do not cause cancer (in the sense of making that conclusion more likely to be true) if a tobacco company were willing to give advertising accounts only to agents who believe that smoking cigarettes does not cause cancer, even though that consideration might provide the rational advertising agent with a reason to have that belief.

So pragmatism seems to be compatible with distinguishing epistemic from nonepistemic reasons, allowing some practical considerations to fall on the epistemic side of this distinction.

1.9 CONCLUDING REMARKS

Despite the clear intuitive distinction we must make between theoretical and practical reasoning, theoretical and practical considerations are rationally intertwined in more than one way. Theoretical reasoning is goal-directed in the sense that goals are relevant to the questions to be considered theoretically and there are practical reasons behind the role of conservatism and simplicity in reasoning.

At present, there is no mathematically elegant account of all aspects of rationality. Formal theories of implication and consistency are possible, but these are only part of the subject. Conservatism, simplicity, and coherence are important aspects of rationality, with explanation, implication, and consistency being relevant to coherence. Our ordinary judgements about rationality and reasonableness are often sensitive to these considerations, but also to strength of will and even fairness.

Logic and probability theory are not directly theories of rationality and reasonableness and, furthermore, it is a misuse of language to say that violations of principles of logic and probability theory are indications of irrationality or unreasonableness. We do not normally consider someone to be 'irrational' or 'unreasonable' simply because of a mistake in arithmetic, or probability theory, or logic. Instead we use the words 'irrational' and 'unreasonable' in a rather different way; for example, for those who refuse to accept 'obvious' inductions, or for those who jump to conclusions on insufficient evidence, or for those who act knowing that they are frustrating their own purposes, or for those who are uncooperative.

These issues are considered further in the next three essays. Essay 2 discusses practical reasoning in more detail. Essay 3 says more about simplicity. Essay 4 takes up the distinction between practical and epistemic reasons for belief.