Introduction: Philosophical Foundations

JONATHAN E. ADLER

1. Reasoned Transitions

Reasoning is a transition in thought, where some beliefs (or thoughts) provide the ground or reason for coming to another. From Jim's beliefs that

(1) Either Bill receives an A or a B on the final.

and

(2) Bill does not receive an A.

he infers that

(3) Bill receives a B.

Assuming that Jim bases his inference on the deductive relation of (1) and (2) to (3), his conclusion is warranted, since the argument is valid. (1) and (2) implies (3), since it is not possible, as contradictory, for (1) and (2) to be true and (3) false. More formally, ϕ is a logical consequence of Γ if and only if there is no interpretation (model) in which all sentences of Γ are true but ϕ is false (Tarski 1983).

Although in reaching (3) Jim comes to a new belief, its information is already entailed by (1) and (2). Unlike deduction, an inductively good argument provides for new beliefs whose information is not already entailed by the beliefs from which it is inferred:

(4) Bill brought his back pack to class every day of the semester.

So, [probably] (5) Bill will bring it to the next class.

The falsity of the conclusion (5) is compatible with the truth of the premises (4). The premises only render the truth of the conclusion more probable (than in their absence). Although this is a good inductive argument, the premises can be true and the conclusion false, so the argument is invalid.

Deductive validity is monotonic: A valid argument cannot be converted into an invalid argument by adding additional premises. But an inductively good argument is nonmonotonic: new premises alone can generate an argument that is not good. If I add to the argument from (4) to (5), a premise that

(4.1) Bill's back pack was stolen,

the conclusion no longer follows.

In either argument, there is a reasoned transition in thought. The person who draws the inference, takes the premises as his reasons to believe the conclusion (or, in the second case, to believe it probable.) By contrast, the transition in thought from the belief that

(6) Joe's cousin drives a BMW.

to

(7) I better call Fred.

is not reasoning because, let's suppose, (6) is merely a cue or stimulus or prompt for the thought that (7) to arise. (6) could not serve as the reason for accepting (believing) (7) as true, as (1) and (2) could for (3). (Another technical use of 'accepting' is for momentary purposes, as, say, when one accepts a supposition for a proof Stalnaker 1987).

Grice (2001) draws the connection between reasons and reasoning by noting that if reason is the faculty which "equips us to recognize and operate with *reasons*" then we should also think of it as the faculty which "empowers us to engage in *reasoning*." Elaborating, he writes

> if reasoning should be characterizable as the occurrence or production of a chain of inferences, and if such chains consist in (sequentially) arriving at conclusions which are derivable from some initial set of

> premises.... of which, therefore, these premises are... reasons, the connection between the two ideas is not accidental. (5)

Grice's 'not accidental' is, presumably, a cautious expression for a conceptual dependence of reasoning on reasons.

Minimally, to have a reason is to have a favorable consideration. However, a reason to do something as in 'my reason to go to the ice cream store is get a sundae' serves to motivate action, whereas a reason to believe does not serve in a motivational role. You can be indifferent to the grade Bill receives, but not, presumably, to the ice cream sundae. Of course, reasons or evidence are typically uncovered through investigation, as when trying to determine the grade Bill receives. But then the motive to investigate obtains independent of the reason to believe. However, in other cases, and much more typically, we acquire evidence that a statement is true and then we come to believe that statement, like it or not. If you overhear Jim affirm (3) that Bill receives a B, then special circumstances aside (e.g., you do not trust Jim), you will come to the corresponding belief, even if you are indifferent to Bill's grade. There is no gap between judging that there are sufficient reasons to believe *p* true and judging (accepting) that *p* is true, nor between judging that *p* is true and believing it.

What is a favorable consideration? Is (1) and (2) a reason to believe (3) because they constitute a mental state or because they constitute facts which serves as the content of that state? If my wanting the sundae is my reason to go to the store, the mental state is the reason. If, instead, what I want to be the case – the fact that I buy apples – is my reason, it supports the truth of the belief's content that I go to the market. ("Belief" suffers a similar ambiguity. Does it refer to an attitude (believing) or to the content of that attitude? We assume that when disambiguation is needed, context will prove adequate.)

Is the reason (as a proposition) a consideration to hold a certain attitude – believing or desiring – or is it a consideration favoring the truth of the content of that attitude (Parfit 2001)? The mother who learns that her son survived a fire in school will be relieved by coming to the belief that he survived, which is then a reason – a consideration in favor – of her taking the attitude of believing it. But that value or utility to her of holding the belief is not a reason that renders it true that her son did survive. In general, it seems that the utility of believing a statement, since it is never a reason for the statement's truth, can never serve as a proper reason to believe. Arguably, though, even if utilities can not bear on what to believe, they may enter with the question of whether to hold a belief rather than not to hold any (Nozick 1993 Ch. III).

2. Belief and Truth

Induction and deduction supply reasons to believe, since each seeks to preserve the truth of its premises, while extending them to new truths acquired as beliefs. Beliefs are the product of reasoning since belief aims at truth. The end result in belief explains why reasoning matters so profoundly to us. We care to get correct conclusions both intrinsically, since that is what having a belief claims, but also, and more obviously, extrinsically. Belief guides actions, and actions are expected to succeed (reach their goal) only if the beliefs that guide them are true. If you want an ice cream sundae immediately and you believe that the only near-by place to purchase one is on the corner of Broadway and 110th St., then you are expected to succeed (to satisfy your desire for the sundae) only if your belief as to its location is true.

Both forms of reasoning or inference aim to discern what is the case, and so aim, figuratively, for the mind to fit the world (e.g., that I come to believe that a sundae is produced at the store just because it is). By contrast, to desire the ice cream sundae, which specifies one's goal in action (to acquire and to eat the ice cream sundae), is to desire the world to conform to the mind. Beliefs and desires have opposite "directions of fit" (Anscombe 1957, Searle 1983).

The fundamental notion of belief is that of "believing that", a characteristic propositional attitude. If Jim believes that Mary is in Alaska, Jim believes the proposition Mary is in Alaska to be true. Propositions are the contents of sentences or statements as expressed on an occasion. The sentence 'I like Krispy Kreme donuts' cannot be true or false as it stands, since the 'I' has no definite reference. But, on an occasion of use, the fixed meaning of 'I' (and similarly for other indexicals like 'you' or 'now') will have their reference determined; the reference of 'I' is the speaker on that occasion (Kaplan 1989). When values for all indexical and similarly context-sensitive terms in an assertion are fixed, the statement expresses an abstract entity of a corresponding form, a proposition. (The prominent features of context are speaker, hearer, location, and time.)

What is it we are claiming of a proposition when we attribute to it truth or falsity? There does not seem to be any difference between

(8) John believes that the proposition that the nearest ice cream store is on Broadway and 110th St. is true.

and

(9) John believes that the nearest ice cream store is on Broadway and 110th St.

Both seem to say the same thing – to be true or false under the same circumstances – suggesting the generalization:

(T) The proposition that p is true if and only if p.

The left-hand side of (T) ("The proposition that p is true") speaks about a proposition. The right-hand side speaks about the world or a fact of the world namely, that the nearest ice cream store is on Broadway and 110th St. The circular appearance does not run deep.

If (T) is correct, there is no further problem about understanding truth than understanding the corresponding proposition. If you understand the proposition that the library is open on Saturday, no special difficulty attends to your understanding the proposition – that the library is open on Saturday – is true (Tarski 1983; Horwich 1990). However, the (T) equivalence does not tell you how to determine or verify or discover whether a proposition is true.

3. Theoretical and Practical Reasoning

Reasoning to how one should act can involve inductive and deductive transitions, but its aim or purpose is distinctive from reasoning whose endpoint is belief:

(10) I want an ice cream sundae.

(11) The closest ice cream store is on Broadway.

(12) There are no barriers to my going there.

So, (13) I should now go to the ice cream store on Broadway.

[Alternatively, (13) I shall/intend to now go . . .]

(10)–(12) constitute good reasons for concluding (believing) that (13) is true. But the ultimate purpose of this reasoning is not to figure out what is the case. Reasoning whose endpoint is belief is referred to as *theoretical reasoning*. Rather, this reasoning (10)–(13) aims to figure out how one should act or *practical reasoning*. The goal is to figure out what one [I] should do (Millgram 2001). As indicated by the alternative reading, (13) should be viewed not just as a judgment as to what is best for me to do, but the actual intention to so act.

Theoretical reasoning aims to answer whether p is the case, not whether I ought to believe it, whereas practical reasoning is concerned to determine what I ought to do. The structure of theoretical reasoning is obscured if its conclusions are taken to be of the form 'I ought to believe p.' What it is best to do is that act which is better than all the alternatives, on the available reasons. But what one can or should believe is only what is genuinely worthy of belief, not what is currently better than the alternatives. (Think here of the difference between poker, where the best hand wins, and rummy, where only the right or proper hand can win Adler 2002).

The end or goal to which practical reasoning is directed is characteristically set by what one wants or desires, expressed in premise (10). (Not that any desire or want specifies a real end or goal – something that you aim to pursue or that even supplies a reason or motive to pursue. You may have a desire to humiliate yourself, which you neither value nor with which you identify.) Practical reasoning aims at figuring out how to go about satisfying a desire, if opportunity permits. When one's wants or desires set a genuine end or goal, motivation to act according to the conclusion's directive is built in. It is unremarkable self-interest to attempt to satisfy one's own ends.

Can one be motivated to act other than internally (from one's wants or desires)? The Humean 'internalist' answers 'no', whereas the Kantian and other 'externalists' answer 'yes.' (Williams this volume). Externalists hold that one can be motivated purely by recognition of a reason (belief) that a rule or principle or duty applies. So, for example, can a child be motivated to visit his grandmother without any desire to do so nor any threat of punishment? Can his recognition that visiting his grandmother is the right thing to do give him a reason to act accordingly, even if he has no internal – desire – motive to do so? Can reason alone, as a source of judgments of truth and falsity, be a source of reasons (motivation) to act?

With slight differences, the internalist answers "no" to these questions, holding that reason is inert. Reason (belief) is only able to guide one to those actions that are likely to

satisfy the motivation that lies elsewhere (in one's desires or wants).

Are one's wants or desires the endpoint to fix one's goal or aims to which practical reasoning is directed? Internalists typically deny that one's ends (or goals) can be rationally altered except on the basis of further desires or wants. One methodological weakness in this instrumentalism is that one's desires are often too unspecific to fix any end. (Richardson 1994; Millgram this volume). Hardly anything is fixed by one's desire to be happily married (to whom?) or to get a good job (which one?). One's ends must be specified to serve as a guide to action, and the specification requires input from beliefs.

Similarly, one's plans need constant updating and modification as they begin to be executed (Bratman 1987). When you learn of a traffic jam further up on the highway, you turn off to the service road. In this way, you fill in your plans, not just modify them. One's plans direct one toward one's goal, but they do so in an open-ended way, leaving room to fill in details and for modifications, as more information is learned.

In theoretical reasoning, motivation is not an ingredient, which is another way to mark its "inertness." Once you judge a conclusion true, based on the reasoning, you thereby believe accordingly, idiosyncratic psychological barriers aside (e.g., distraction). Belief is in one way passive and not subject to choice: Think of all the beliefs you pick up on the way to your morning commuter train to which you are completely indifferent, for example, that your new neighbor is wearing a green jacket today. No motivation is necessary for belief to respond to a convincing argument. It is a heard contradiction, discussed further below, to affirm a statement of the form "*p* is true, but I don't believe it" ("Moore's Paradox").

The objective of theoretical reasoning is to relieve doubt or to satisfy curiosity or to diminish puzzlement by achieving corresponding beliefs, whereas the objective of practical reasoning is to secure the means to realize one's ends. Because practical reasoning is directed toward action it is overtly constrained by time and resources – its objective is to discover which option is best, all available things considered. But the objective of theoretical reasoning is not merely to discover what proposition (option) is best supported by one's available evidence, but what is correct (true). Consequently, to draw a conclusion in theoretical reasoning requires the claim not just that one's evidence is the total relevant evidence available, but that the evidence is representative, rather than a skewed sample. It follows further that our limits in gathering and assessing evidence contours theoretical reasoning, but, like utilities, our limits cannot play an overt role in drawing a conclusion. You should not – and, perhaps, cannot – believe a conclusion because it is best supported by the evidence so far and that you do not have more time to examine further evidence. (The problems raised here are for assimilating theoretical to practical reasoning. For a discussion of the assimilation of practical to theoretical reasoning, see Velleman 2000.)

4. Theoretical Reasoning: Limits, Closure, and Belief-Revision

Our limits restrict the resources and time to devote to empirical search, testing, and inquiry, as well as to the inferences worth carrying out. The valid and sound argument from "Trump is rich" to "Trump is rich or cousin Harry is in Jamaica" yields no new worthwhile information. Endless such trivial consequences (e.g., *p*, *p* or *q*, *p* or *q* or *r*, . . . ; *p*, *p* & *p*, *p* & *p* & *p* . . .)[1] can be so generated, which will just "clutter" one's memory as explicit beliefs (Harman 1986; Sperber and Wilson 1986). Also, if one "loses" or forgets the origination of the disjunctive belief in the belief that Trump is rich, one will mislead oneself on attending to it that one has special reason to believe Harry is in Jamaica or that it bears a significant connection to the other alternative that Trump is rich.

Theoretical reasoning involves revising beliefs we already hold. Rules of standard logic or implication, however, do not (Harman 1986). Jane believes that if she attends Yale, she'll become an atheist. She believes that she will attend Yale. If she reasons by the impeccable rule of Modus Ponens (MP: *p* and if *p*, *q* implies *q*), she concludes that she will become an atheist. But although she now has a reason to believe that conclusion, logic does not decide that she will or should believe it. Once Jane becomes aware of that conclusion, she also becomes aware of other beliefs, which deny that she will ever be an atheist. Instead of drawing the MP conclusion, Jane ceases to believe the conditional, which served as her main premise. Reasoning that results in modification of beliefs of one's own may be dubbed "self-reductios" (ad absurdum).

Examples such as the previous one show how from deduction we can learn something new about the content of our beliefs, even though, in a figurative way of speaking, deduction only

renders explicit information already in the premises. Briefly, our beliefs are not closed under deduction. (Similar, but distinguishable, worries attend to the requirement that our beliefs at a given time be consistent. The worries are different for failures of closure or consistency that the agent does not recognize and those cases in which the agent does recognize the failure. The latter cases generate more forceful conceptual friction with the concept of belief. For recent treatment of these logical requirements on belief, see Christensen 2004.) One's beliefs are closed just in case if one believes p and p implies q, one believes q. None of us mortals have bodies of beliefs that are deductively, let alone inductively, closed. There are complex tautologies or logical equivalents to what we believe, which we will not believe and may even disbelieve. The failure of deductive closure for belief is a facet again of our limits, including our limited grasp of our own beliefs, our lack of omniscience, and our "inability" to perceive the future. If Socrates believes that no one does wrong knowingly, does Socrates believe that Richard Nixon did no wrong knowingly? (For examples and critical reflection's see Stalnaker 1987: Ch. 5.) Implications or deductions from one's beliefs can yield surprising conclusions. Well before the discovery of penicillin by Fleming, biologists knew that molds cause clear spots in bacteria cultures, and they knew that a clear spot indicates no bacterial growth. Yet, they did not come to realize, or even to hypothesize, that molds release an antibacterial agent. The observations did not render salient the disparate beliefs and place focus on them together (Cherniak 1986).

So, putting aside closure imposed as an idealization for specialized purposes (Hintikka 1962), one can believe, or even know, p, and p imply q, without believing or knowing q. Similarly, it can be the case that a = b and that one believes (and even knows) that a is F, without one believing that b is F, though the embedded argument is valid. So, for example, Lois Lane may know that

(14) Superman flies.

In the tale, it is true that

(15) Superman is Clark Kent.

But Lois Lane does not know (and actually believes false) that

(16) Clark Kent flies.

The fault lies with a lack of knowledge of the middle step – (15). Knowledge or belief is "opaque" – in "S believes that a is F" the position of "a" is not purely referential. (Opacity intrudes on what counts as a reason: If Lois Lane wants to marry only a man who flies, does she have reason to marry Clark Kent?) Consequently, substitution of arbitrary coreferential terms is not truth-preserving. Within the scope of Lois Lane's beliefs or knowledge, "Superman" in (14) does not simply refer to an object (Superman), but to that object as understood by Lois Lane. (The problem originates with Frege [1970]. An alternative account holds that the substitution does go through. The assumption as to how the person [Lois Lane] thinks of the name is only pragmatic. A parallel worry applies to the distinction between attributive and referential meanings of a term [Kripke 1977]. To appreciate this alternative reading substitute for the names a pure pointing device like "this" or "that.")

A much-discussed example takes the problem of closure a step further, because it holds that knowledge is not closed even when the person knows the "middle" step – the relevant implication (Dretske 1970; Nozick 1981). Assume that Tony is looking at an animal behind a cage marked "zebra," which looks like a zebra. Barring any weird circumstances, we would say that Tony knows that

(14) The animal I am looking at is a zebra.

Let's now grant that Tony also knows the implication that

(15) If the animal I am looking at is a zebra, then it is not a mule cleverly disguised to look like a zebra.

(14) and (15) imply

(16) The animal I am looking at is not a mule cleverly disguised to look like a zebra.

Still, we are reluctant to attribute knowledge of (16) to Tony. Those who oppose closure reason that Tony has never checked that the animal he is looking at is not such a cleverly disguised mule. Tony is simply looking at the animal from outside the cage. These theorists reject:

> (Epistemic Closure EC) If X knows that p and X knows that p implies q, then X knows that q.

Arguably, this is the principle licensing Descartes' famed sceptical argument: If you know that you are in your office and you know

that if you are in your office, you are not just *dreaming* it, then you know you are not just dreaming it. But you do not know that you are not dreaming it. So you do not know you are in your office.

The rejection of EC fits the previous zebra example, answers Descartes' sceptical argument, and it is explained as due to our not checking on all the implications of propositions that we know. The rejection also follows from analyzing knowledge as involving satisfaction of the following subjunctive or counterfactual conditional:

> (Tracking Knowledge TK) Were *p* false, S would not believe *p*.

The most likely (or nearest) way for it to be false that you are in your office is for you to be somewhere else, like your kitchen. If so, you would clearly recognize where you were in the other room. Consequently, you would satisfy (TK), because you would not believe that you were in your office. (TK) does not then support (EC). (Contextualists, whose views we return to below, hold that you do know that you are not dreaming, when you are in an ordinary setting. However, when Descartes or a skeptic mentions the possibility that you are dreaming, they alter the context or standards for knowing. Only then you do not know that you are not dreaming. But, in that case and compatible with EC, you do not know that you are in your office either.)

Despite these advantages, the dominant view is that EC cannot be rejected, since deductive implication preserves truth. What better way to know the truth of a proposition but by deducing it from a proposition one does know? (For overview, see Luper 2006). Without pursuing this line, it's worth noting that sometimes (non-trivial) deductions seem not be a way to advance knowledge. From the evidence of (17),

(17) The Smiths are making an extravagant wedding for their daughter.,

(18) is concluded:

(18) The Smiths are wealthy.

From (18), (19) follows:

(19) In making the extravagant wedding, the Smiths are not just appearing to be wealthy.

Assume that you are in a discussion with someone who disputes whether the Smiths are really wealthy. Although (18) implies (19), it seems to *beg the question* in this context to use (18) as a reason to believe (19). (17) can only provide evidence for (18) if (19) is assumed or presupposed. But (19) is in dispute. If it is presupposed in treating (17) as evidence for (18), then the warrant or support that (17) lends to (18) does not *transmit* to the conclusion (19) (Wright 2000).

5. Belief-Revision, Holism, and the Quine–Duhem Thesis

If a corpus of beliefs is not closed for the reasons suggested, it is likely to be inconsistent. If there are serious implications of one's beliefs that one fails to believe, one is likely to acquire the contrary of some of those beliefs without recognizing the incompatibility. Here's a very ordinary illustration of Lewis's:

> I used to think that Nassau Street ran roughly east-west; that the railroad nearby ran roughly north-south; and that the two were roughly parallel. (1982, 436)

Once these beliefs are brought together with the evident tacit belief, Lewis recognizes that the set of beliefs {Nassau Street ran roughly east-west; the railroad nearby ran roughly north-south; Nassau Street and the railroad nearby are roughly parallel; if one path is east-west and another is north-south, they are not parallel} is inconsistent. They cannot be simultaneously true. Once Lewis recognizes the inconsistency, he can no longer hold on to all these beliefs ("I used to think..."). The question that he now confronts – the question of belief-revision – is how he should restore consistency.

Rejecting any one or more of the members of the inconsistent set will restore consistency. Quine (1980) argued that selection to restore consistency depends on extralogical considerations. He made these claims in developing his criticism of the "dogma of empiricism" that statements (hypotheses) can be tested in isolation. Instead, he put forth what is referred to as the "Quine–Duhem Thesis" (Duhem 1954) that hypotheses are never tested in isolation. Hypotheses (or theories) do not entail any observational predictions by themselves. To derive predictions that serve to test a hypothesis, assumptions are required that crucial terms are not empty and that conditions are normal. Newton's enormously successful theory of gravity and mechanics erred in its (pre-1846) prediction of the orbit of Uranus. But Uranus's deviation was treated as an anomaly, rather than

a falsification, because the theory made substantial assumptions about the operative gravitational forces. In the discovery of Neptune, some of those assumptions were abandoned, rather than the Newtonian theory itself. In general, when a well-regarded hypothesis fails, we do not immediately conclude that the hypothesis is false, as the traditional view implies, rather than that some of the conditions assumed normal – the auxilliary assumptions – failed.

The hypothetico-deductive model incorporating the Quine–Duhem thesis is represented schematically as

H and Auxilliary Assumptions (AA) imply O.

If O, then (H and AA) are confirmed.

If not O, then (H or AA) fails.

The latter is the crucial result because consistency does not demand the falsity of H.

The problem of what extralogical principles to apply to belief-revision has generated numerous investigations and constructions of logics of belief revision (Hanson 2006; for an introduction to a computational approach to belief reasoning and revision, see Pollock and Cruz 1999: Ch. 7). A central proposal is that belief revision should be conservative. One revises one's beliefs so that rejection or modification is minimal. You all-out believe that Skinner wrote *Walden II* and that Chaucer wrote the *Canterbury Tales*. But if you discovered that one of these is wrong, you would sooner surrender one rather than both and the latter, rather than the former, which is attested to by a greater variety of good sources. To surrender the belief about Skinner's authorship would require surrendering – nonconservatively – much more information than surrendering the latter.

But conservatism cannot stand alone. Lewis cannot just decide to give up the belief that Nassau Street ran roughly east-west and keep the one that the railroad nearby ran roughly north-south, although that surrenders only one, rather than more, among equally contentful, incompatible, beliefs. Merely his deciding would not be a sufficient reason that the belief retained is true.

Other principles of belief revision include prominently simplicity and coherence. The more a belief coheres, fits, or is explanatorily connected, with others, the more resistant it should be to modification. But some conflictual beliefs may be surrendered (in strain with conservatism) to increase coherence. You believe that ten-year-old Jim is in good health, that he will be in the tennis tournament tomorrow, and that he will meet you for lunch at noon. You learn that he is not at school today and that his mother is not at her office. You infer that Jim is sick. That best explains the latter two beliefs – unifies them in an explanatory nexus. But as a consequence you surrender other beliefs about Jim (e.g., that he will be at tennis practice later in the afternoon).

Coherence is an internal requirement on one's beliefs. But our belief corpus improves by external input, especially through our senses. The improvement is not just in the addition of new beliefs picked up as we navigate our environment. Perception and other sensory mechanisms provide for ongoing self-correction, which is a hallmark of scientific method (Sellars 1963). If you believe that Lisa is in Alaska and you see her car at the local diner, you surrender – and so correct – your belief. Normally, beliefs operate as a filter on perceptual judgment. If Lisa drives a blue Ford and that was all that was in your perceptual field as you approached the diner, then if you noticed the car, you would not think of it as hers, given that you believe that she is in Alaska.

However, there must be limits to this filtering role, otherwise our beliefs would not be subject to correction. Once you see the car more closely, and observe the familiar dent on the hood, you are compelled to notice that it is Lisa's blue Ford. Your belief is revised. The new perceptual information nullifies your prior belief, evidence that perception can succeed as a self-corrective on reasoning only if it has some independence of operation from belief and reasoning (from one's "central systems," Fodor 1983).

If the formation of perceptual beliefs always had to be first checked (for veracity) by way of one's corpus of beliefs, it would be subject to the "dogmatism paradox." If you know that Lisa is in Alaska, why should you even acknowledge as putative undermining evidence that it is her car at the Brooklyn diner? Shouldn't you rather judge that, say, her husband must have taken her car, because, if you know that she is in Alaska, shouldn't you know that putative evidence against it must be mistaken or misleading? [Think of the tautology: p–>((q–>~p) – >~q).]

However, there is something deviant about the conditional "If there is evidence against my knowledge (that Lisa is in Alaska), then that evidence is mistaken or misleading." It does not

seem to be open to modus ponens, just as the following is not:

> If my wife cheated on me, I would never know. (Harman 1973; Ginet 1980; Stalnaker 1987)

Were I to discover that my wife cheated on me, I would reject the conditional (or its antecedent), rather than conclude that I would never know that she cheated on me. Similarly, the previous conditional of the dogmatism paradox is to be rejected when the undermining evidence is obtained, rather than rejecting the evidence as misleading or false.

6. Deductive Rules and Deviant Logics

Usually, of course, conclusions drawn from one's beliefs simply form new beliefs. But often conclusions are drawn that cast doubt back on the premises (beliefs) or the inferential transitions from which those conclusions are drawn.

In drawing out conclusions, what rules should be used? Although in the case of induction and especially deduction a core of rules and results are well established, disputes abound about their scope and other putative rules are flat-out contested. After considering some of these rules and disputes, we briefly turn to how the rules are to be justified and selected.

Standard or classical logic is first-order quantification or predicate logic – the logic of the truth-functions ("and," "or," "not," "if, then") and the quantifiers ("For every . . . ," "For at least one . . . ") The "first-order" implies that the variables of quantification take as values objects or individuals, not names, predicates, propositions, or properties. First-order logic (including identity) is sound: no proof (a syntactic notion) will take one from truths to falsehoods. Every proof corresponds to a valid argument, a semantic notion. But also and more distinctively, first-order logic is complete (every logical truth or valid argument is provable). Once second-order quantification is admitted, particularly to embrace set theory, the extended logic is no longer complete.

Unlike additions to standard logic, as in adding axioms for necessity and possibility to logic, deviant logics deny some basic logical law. Quine's "holism" opened a door to defend deviation from classical logic, which Quine (1970) attempted to quickly shut. The holistic assumption that justification for logical laws, like the justification of empirical claims, is sensitive to the whole body of beliefs provides an opening for arguing that a logical law is to be rejected because removing it from one's corpus of beliefs increases coherence. But although Quine's opened this door as a theoretical possibility, he argued that the standard logical laws are too useful or indispensable to the progress of science for abandonment.

Additionally, Quine argued that when you try to deny a logical law like the law of noncontradiction (i.e., $\sim(p \ \& \sim p)$, your "&" and "$\sim$" no longer translate "and" and "not," as intended. These operators are fully specified by the truth-tables and the implied laws. They are defined implicitly by their roles in these assignments and laws. (However, we know that unless restraints are imposed on implicit definitions, specifically, that they introduce no new theorems ["conservative"], the introduction of new connectives can generate crazy rules, ones from which anything can be deduced; Prior 1960; Belnap 1962). If someone infers from an utterance of "A or B" to "A," we can be sure his "A or B" is not the English disjunction. The deviant logician's predicament is that "when he tries to deny the doctrine he only changes the subject" (Quine 1970), p. 81.

Quine's argument opposes the plausible claim that the denial of an operator in some inferential roles is compatible with its playing the appropriate roles in other inferences, sufficiently so as to remain a viable candidate for capturing the basic meaning. Even the denial that all contradictions are false, allows for preserving a good deal of classical logic with suitable adjustments. In classical logic, every sentence follows from a contradiction. However, this trivialization can be excised by denying the rules from p to $p \vee q$ (weakening); and from $p \vee q$ and $\sim p$ to q (disjunctive syllogism)). In the former case, the denial is independently motivated by the lack of relevance in subject matter of the premise to the disjunct q. With a number of other adjustments a major portion of basic logic remains in tact, and the resulting logical system can be sound and complete (Priest 1998).

Although most of the alleged examples of contradictions that are candidates for the ascription of truth are rarified, one of them is a resource from which a number of deviant logics draw strength, and it resonates with everyday reasoning. The resource is in the phenomena of vagueness. The vagueness of a term is that while it sorts objects into those to which the term applies and those to which it does not, it leaves undecided many objects. When a teenager's room has no clothes on the floor and dirty dishes

have been removed, but it has not been dusted or swept, it is indeterminate, let's suppose, whether it is clean or not. A defender of the view that there can be true contradictions might say that the room is both clean and not clean. Others may deny the law of excluded middle: that either the room is clean or it is not clean.

A contextualist confronted by an assertion such as "John's room is clean" will respond that for everyday purposes, it is enough that his clothes are off the floor and that the dirty dishes have been removed. But if John develops asthma, you will not count the room as clean until a careful dusting is complete. The proposition expressed by the assertion is false in that context. Outside of any contextual specification, there is no assigning it a truth-value (or assigning it additional truth-values) (Lewis 1983). Contextualism explains how one utterance of "John's room is clean" can be true and another false, when there is no change in John's room. Because contextual variations include variations in the importance of the matter, contextualism also makes sense of why when you inquire as to the truth of a hypothesis, you are bound to investigate harder in one context (where the costs of error are greater) than in another.

Contextualism, however, is a minority view in how to handle the sceptical implications of vagueness that originates from the "sorites paradox." In its historical form, the sorites is presented as the "paradox of the heap" (Sainsbury 1988; this volume). If you have a heap of sand, and you subtract one grain, you still have a heap (one grain cannot make a significant difference). The judgment suggests a principle: if you now subtract one grain from the previous heap, you still have a heap. But repeated applications leaves you with a couple of grains that you are committed to taking to be a heap, even though they obviously do not make a heap.

Intuitionists respond that when there are too few grains in a pile to clearly be a heap and too many grains to clearly not be a heap, it is not true that either that pile is a heap or that it is not a heap, contrary to the law of excluded middle. And if it not clearly either, the failure to not be a heap does not imply that it is a heap, contrary to the law of double-negation.

The sorites is derivable via the impeccable principle of mathematical induction: R_o is the base case with a certain property P (e.g., That large collection of grains of sand is a heap). And if R_n has P, there is some fraction of it (e.g., one grain of sand) such that if we decrease R_n by that amount, then what results – Rn-1 – still is P (e.g., that 1-less grain collection is a heap). Then any lesser R_i has P (e.g., every collection of grains of sand less than the base case is a heap including a one-grain collection). Alternatively, the paradox can be presented simply as a string of MP arguments, each one yielding a decrement from the previous. Either way, the nonnoticeable difference for any sized heap – the decrement between R_n and R_{n-1} – becomes a marked difference after enough applications.

The sorites paradox is particularly wrenching because it seems to arise merely from the vagueness of terms, which holds of most terms. Most, if not all (nonartificial) terms leave undecided an unlimited range of cases for example, "blue," "happy," "short," "table," "flat," "rich," "child." At what moment does childhood end? The exception would be contrived cases where an exact specification is provided: We could define a U.S. adult citizen as rich* just in case his total wealth is $484,234.04 or higher. However, to attempt exact replacements for our vague terms would fail to preserve their value or usefulness. (On utility considerations for vagueness, see Parikh 1994.) The contrived precision would require a sharp break in judgments, where a gradation of responses is appropriate (between, e.g., a person who is rich and one who is very well off financially). The vagueness of a term reflects its "tolerance" for certain tiny alterations, which can mount up to significant alterations.

But is this insensitivity in the term or is it merely because of the limits of our discriminatory powers, which are foisted on to the term? "Epistemicists" favor the latter, which allows them to avoid offending against standard logic. They hold that there is an exact boundary for vague terms, but it is unknowable (Williamson 1994; Sorensen 2001). (The previous question suggests another: Is it reality – heaps themselves – that is vague or how we describe it, e.g., some collections of grains of sand are described as "heaps"?)

Among the numerous attempts to solve the problem the dominant view preserves almost all of standard logic, allowing only for truth-value gaps. A "supervaluationist" observes that within the indeterminate cases, we are free, as far as a consistent assignment of truth-values, to decide them as serves our purposes, as contextualists claim, too (Fine 1975). Some will treat a U.S. citizen with total wealth of $325,683.03 as rich and others not (for purposes of assigning, say, an estate tax), because this amount is clearly between the definitely not rich and the definitely rich. When we so decide cases we

provide "sharpenings." However, on every sharpening "That citizen is either rich or not rich" will be true, so supervaluationists can accept the logical law of excluded middle (although neither disjunct may be definitely true). Consequently, on a supervaluationist view, a conditional sentence – the second step in the sorites paradox (e.g., if that citizen is rich with $325,683.03 then he is rich with $325,683.02) will not come out true for every sharpening of it. There is a sharpening under which the antecedent is true, but the consequent false, so that the conditional is false. Because supervaluationism does not reject logical laws and it does not require powers in our language that supercede our own, it has the advantage of providing for a conservative response to the sorites.

But is the supervaluationist right, to return to the earlier example, that '"either that room is clean or not" is definitely true, when the room is clearly a borderline case? Worries like this incline others, although far fewer, to take the route of treating the initial reaction that neither alternative is true at face value. We can say only that John's room is clean to a certain degree, or to a higher degree than others. The error on this probabilistic approach is to contrive to derive absolute judgments from matters of degree.

7. Ordinary Language Challenges to Logic and the Conversationalist Response

A very different source of doubts about logical reasoning as standard first-order logic derives from alleged deviations from ordinary language. Numerous patterns of inference of ordinary language, as well as straightforward readings of complex statements, prima facie do not obey the rules governing deductive logic or the logical operators. Some examples:

(20) John goes drinking and John gets arrested.

(21) John gets arrested and John goes drinking.

If "and" is the "&" of formal logic, it is symmetric, so (20) and (21) should be equivalent. Yet, they do not seem to mean the same, (21) does not follow from (20) (or conversely). Another:

(22) John will order either pasta or steak, but he orders pasta.

So (23) John does not order steak.

The inference seems valid, but fails on the truth-table analysis of "v" (inclusive "or"). Finally for an example using a conditional to which we devote the next section:

(24) If you tutor me in logic, I'll pay you $50.

So (25) if you don't tutor me, I won't.

The conclusion seems to follow. However, the straightforward translation of it into logic yields a fallacious form, one that appears valid, but that isn't.

One reaction to such discrepancies is: so much the worse for ordinary language. It requires formal regimentation to be a satisfactory medium of reasoned argument. The opposed reaction is: so much the worse for logic's claim to provide a systematic analysis of ordinary reasoning.

The most profound reaction is that of H. P. Grice's (1989). His account of conversational reasoning opposes both previous ones. Grice's claim is that the logic of ordinary language is already that of formal logic. However, we impose, without recognition of the imposition, assumptions or expectations on ordinary language because we treat the sentences as assertions or other contributions to a conversation. In (22)–(23), we assume that John could not order both pasta and steak, given our knowledge about eating. The inference from (20) to (21) (or their equivalents) stands. However, the speaker exploits the listing of conjuncts, which is mutual with the hearer, as implicating an ordering (in time). Grice's account explains these deviations without positing an ambiguity in the relevant logical constant (e.g., "and," "and then").

Conversation or social communicational exchanges are facilitated by shared or mutual assumptions that are not part of what is said or its logical implications, but which nevertheless are invited, given the common goals of the cooperative exchange. The fundamental "maxim" (and so expectation or presumption) is that the speaker intends to cooperate to advance the purposes of the conversational exchange. The Cooperative Principle (cp) includes subsidiary maxims. The speaker intends his contribution to be informative, warranted, relevant, and well formed (for brevity, style, politeness, and comprehension). This package of maxims under the cp Grice thought to be justified as principles for rational cooperative arrangements for beneficial ends (of transferring information).

What we mean to communicate typically goes beyond what is said, although calculated on the basis of what is said:

H (hearer): Are you going to Jeff's party?
S (speaker): I have an exam the next day.

S's stated response has nothing to do with H's question. But H presumes that S is following the cp, and that is best explained if S meant to communicate a negative answer, as well as supplying H with further information (as to her reason). S conversationally *implicates*, in Grice's term, that she will not attend. If, however, the assertion's only function was to yield this implicature, then the burden on the hearer of drawing the inference, and the risk that he would not succeed, is not worthwhile: Why not answer the question directly? Consequently, assuming that the cp is in force, the speaker must intend to communicate further thoughts (beliefs) – that the studying is his reason for not attending, and that the activities are incompatible. She efficiently conveys much more information than she would by a mere "no." (Optimization of costs and inferences is central to the pragmatics of Sperber and Wilson 1986).

The technical term "implicature" is meant to draw a comparison with logical implication, yet to distinguish them. The best, although still imperfect, test for implicatures, as contrasted to logical or semantical implications, is *cancellation*. This speaker could say without contradiction: "I have an exam on the next day, but maybe I'll go to the party anyhow." The second clause cancels the implicature of the first. In (20)–(21), the ordering is arguably not part of the meaning because we can append to either, without contradiction, a cancellation "... but maybe not in that order."

Cancellation provides a way to establish that the fundamental speech-act of assertion does not merely conversationally implicate that the speaker believes what he asserts. For, again, it is contradictory to assert any statement of the form "*p*, but I do not believe that *p*," which is the basic form of Moore's Paradox. The paradox is that this sentence is consistent: It can be raining (*p*), say, but I not believe it. Yet, it is inconsistent to assert. The impossibility of canceling the implicature leads Grice to explain the "heard" contradiction as stemming from assertion expressing belief, not merely implicating it.

Because the expectation generated by the maxims are mutually known to speakers and hearers, speakers can exploit them to communicate better. Speakers can overtly violate a maxim as with metaphor or irony. In Plato's *Euthyphro*, when Socrates tells Euthyphro "I should be your student, Euthyphro," what he asserted is recognized as blatantly unwarranted by his audience (and it is expected to be so recognized by the speaker), so that Socrates implicates the opposite.

Grice characterizes when *p* conversationally implicates *q* as follows:

> He [the speaker] has said that *p*; there is no reason to suppose that he is not observing the maxims, or at least the Cooperative Principle; he could not be doing this unless he thought that *q*; he knows (and knows that I know that he knows) that I can see that the supposition that he thinks that *q* is required ... he intends me to think ... that *q*; and so he has implicated that *q*. (31)

These chained inferences require the speaker to assume not just that the hearer knows the maxims under the cp, but that the hearer can work out or calculate the implicature, based on his presumption that the speaker complies with the cp.

Grice's essays on conversational implicatures have given rise to a vast amount of extremely fruitful research on pragmatics, spanning a number of disciplines.

8. Conditionals and Conversation

Of all the originally intended applications of Grice's pragmatics to ordinary language inference, the most promising and the most difficult is to the indicative conditional, such as "If John's car is in his driveway, then he is at home."[2] The wide disparities between the ordinary indicative conditional (If A, B) and the material conditional ($\supset$ or 'hook') of propositional logic are well known. Aside from the material conditional not requiring any overlap in content between antecedent (A) and consequent (B), counterintuitive results follow when the antecedent is false, because then the material conditional is true. (The only secure row in the truth-table for the material conditional is falsity, when antecedent is true and consequent false.) For one problem case: Assume it is false that Jones fails the final. Then it will be true for the material conditional that if Jones fails the final, she will be overjoyed (and also, of course, that if Jones fails the final, she will not be overjoyed).

Grice's suggestion for the alleged failings of the material conditional is that we confuse whether the conditional is true with whether it is

assertible (without misleading). I know that if it is false that Jones fails the final, it would be misleading to assert the conditional. The conditional is noticeably weaker than the relevant information I could assert instead with no more burden on the hearer namely, Jones will not fail the final. (Compare to: either Jones does not fail the final or she will be overjoyed, which is logically equivalent to the material conditional). So the hearer will take the speaker to have (falsely) implicated that she does not know whether Jones fails the final. To explain away the irrelevance problem, a Gricean proposal is that asserting two sentences close to one another would be confusing (disorderly), unless they enjoy some common relation to the informational purposes of the exchange. In the case of the conditional, the relation would be some reason or ground (expressed in the antecedent) for the consequent.

Jackson (1987) developed a Gricean approach. With qualifications, the conditional is the material conditonal. But it is assertible only when the consequent remains highly probable (is "robust") in the event that the antecedent turns out to be true. That is, not only should the conditional be highly probable in itself, but it should also be highly probable *given* that the antecedent is true. The reason for this is that upon learning that the antecedent is true, we want to be able to proceed with such inferences as modus ponens, rather than having to withdraw the (material) conditional. Although it may be true that if Jones fails the final, she will be overjoyed, if it is true only because she will not fail, it is unassertible, since it is not robust with respect to its antecedent.

Grice's proposal, particularly as developed by Jackson, contains insights that all draw on. Tom believes that his son Joe will get his driver's license, regardless of whether he takes a driver's ed course. So it will be misleading for him to assert, even if circumstances are otherwise appropriate, "If Joe takes the driver's ed class, he will get his driver's license," although Tom does believe this. Still, the objections to the Gricean analysis are formidable for thought, not just its home territory of conversation. If you believe it is false that Jones fails the final, you may still disbelieve (and not just refrain from asserting) that if she fails, she will be overjoyed. Or, because the negation of a material conditional implies the antecedent and negation of the consequent, I will disbelieve (believe false) that if I get arrested tonight, I'll go dancing in the morning, while disbelieving, as well, that I will be arrested tonight.

A widely adopted proposal denies Grice's argument that the conditional is actually truth-functional. Instead, the "if" directs a supposition of the antecedent. (Bennett 2003 for a survey with references to the main contributors.) If, given that supposition, the consequent holds (or it is highly probable) within one's corpus of beliefs when the corpus is minimally altered to accommodate to the antecedent, the conditional is assertible. (An alternative formulation is via the similarity of possible worlds.) In short, the assertibility (warrant; acceptability) of "if A, B" equals that of pr(B/A), where this is the subjective conditional probability of B given A, determined in accord with the suggested procedure.

The 'suppositional' proposal follows upon the major discovery of Lewis's (1986b) that the simple equation of pr(A–>B) and pr(B/A) (when pr(A)>0) fails. Conditional probabilities cannot, without trivialization, be the probability of the truth of a (conditional) proposition. Consequently, one should not view "if A, B" as a truth-bearing proposition, but as an evaluative procedure whereby the proposition A is supposed and B is evaluated on that supposition. The result of that evaluation is not, however, a proposition that is probably true, when B is probably true, given A.

There have been extensive developments of this framework, and the discovery of a number of sticking points such as how to handle embedded conditionals and cases in which it seems that two persons could each have sufficient reasons to accept conflicting conditionals A–>B; A–>~B. (McGee 1985; Gibbard 1981)

9. Foundational Problems of Induction

Gricean implicatures are inductive inferences or inferences to the best explanation (what the speaker meant as the best explanation of why he asserted what he did). Aside from a dominant commitment to the probability calculus, there is no analogue for inductive logic of the basic laws or rules of first-order logic to which deviant inductive logics may dissent. Even so basic a rule as the "straight rule" (if m/n As are Bs infer that the probability that the next A will be a B is m/n) has been questioned (doesn't it matter under what conditions the As were discovered – a single or varied sample?).

Although it seems evident that positive inductive evidence for a hypothesis provides support for it, does that support actually provide reasons to believe that its claims will continue to hold (in the future)? A negative answer

was offered by Hume, at the launching point for investigations of induction. Hume writes

> We have said, that all arguments concerning existence are founded on the relation of cause and effect; that our knowledge of that relation is derived entirely from experience; and that all our experimental conclusions proceed upon the supposition, that the future will be conformable to the past. To endeavour, therefore, the proof of this last supposition by probable arguments, or arguments regarding existence, must be evidently going in a circle, and taking that for granted, which is the very point in question. (Hume 1977: 23)

Nondemonstrative arguments about an unobserved (future) event only work if we assume that "the future will be conformable to the past," which is referred to as the assumption of the uniformity of nature (UN) – the future will be like the past (at least in respect of the regularity involved). So the argument should really be:

In the past, Fs have been followed by Gs (and never by non-Gs).

This present case is an F.

UN: Nature is uniform (at least in regard to Fs followed by Gs).

So, the present case of an F will be followed by a G.

The cogency of this argument turns on the uniformity supposition. Its defense can not be, again, by demonstrative argument, because it is possible that the future is not like the past in this or any other respect. The laws could break down. ("It's possible," in Hume's weak sense of possibility as logical consistency, that the sun does not rise tomorrow.) So then we need a parallel argument for uniformity:

In the past, nature has been uniform (at least in regard to Fs followed by Gs).

The present case is an instance of that uniformity (of an F).

So, the present case will be followed by a continuation of the uniformity (a G will follow).

Because the premises are all about the past and the conclusion is about the future, what is needed is a past-future linking assumption:

In the past, nature has been uniform (at least in regard to Fs followed by Gs).

The present case is an instance of that uniformity (of an F).

Nature is uniform (at least in regard to the uniformity of Fs followed by Gs).

So, the present case will be followed by a continuation of the uniformity (a G will follow).

But the premise affirmed is effectively the conclusion itself or a proposition that implies it. The reasoning is circular in a way that would be found for any inductive or nondemonstrative argument, since they all require the uniformity assumption.[3] In brief, even if induction has worked in the past to infer that it will continue to work in the future would itself presuppose the validity of inductive inference, which is what we were supposed to prove.

Karl Popper (1959) champions Hume's inductive scepticism, stressing that Hume's doubts or scepticism apply even if the inductive conclusion is to follow only with high probability. Deduction is really the only kind of reasoning. Even if no amount of evidence can increase the inductive probability of any statement, one negative instance is enough to falsify it conclusively.

Popper's falsificationism claims that scientists should aim only to falsify, not confirm, hypotheses and so they should express hypotheses in as strong a form as possible. Falsificationism runs up against many technical hurdles. Most relevant is a nontechnical problem. Because a strict falsificationist denies any inductive inference, he cannot make sense of the relation between finding positive evidence and rational increases in confidence.

Other responses to Hume's problem include the proposal to vindicate induction, showing that even if induction cannot be justified, it is the best or the only method that works if anything does (i.e., if nature is uniform) (Reichenbach 1961) and the "Oxford" position that it is a conceptual truth that an empirical statement is justified by positive evidence (Strawson 1952). The latter, however, seems to only push the problem back to determining what counts as positive evidence. Bayesian views, discussed below, hold that opinions or subjective judgments are free for the asking. Conformity with the axioms of probability is the only constraint. But with new evidence our confidence in those opinions and judgments, as hypotheses, are rationally alterable according to a probabilistic analogue of logical consistency and a rule for learning (or updating values). Together these equate the new

probability with the old conditional probability of a hypothesis on that new evidence. The justification of inductive judgments should be viewed diachronically, not synchronically.

We know that overwhelmingly inductive inferences do work. They must or else our success in action, and that of other animals, makes no sense, since induction is our guide to forming expectations and to the future. This is not an answer to Hume's sceptical argument, but only a way to clarify its scope. Even if we cannot demonstrate that inductive arguments must be reliable, it is enough for our confidence in induction that it regularly succeeds. The possibility that the future is unlike the past is not often realized and the world is by-and-large lawful, even if neither observation can be massaged into a justification of induction.

10. Qualitative Confirmation and Its Paradoxes

Although Hume's analysis of induction is the background to the development of an inductive logic, its nearer roots are the project of articulating a logic of scientific method. The project takes off from the failure of Positivist or Logical Empiricist attempts to provide a criterion to demarcate science from metaphysics or nonsense, or, less tendentiously, nonscience. Because scientific hypotheses are typically in the form of generalizations – for example, "F = ma" – whose scope is unlimited, and because our evidence for any hypothesis at any time is finite, hypotheses could not be established deductively with premises that only report the evidence.

Our hypotheses are highly underdetermined by the evidence in their favor. Developing a confirmation theory or inductive logic was aimed at the fundamentals of epistemology – to provide standards of rational belief. Initially, the attempt was to propose and examine rules of qualitative confirmation, whose findings were expected to be incorporated into a quantitative account (Hempel 1965).

An attractive starting point is that a generalization of the form "All As are Bs" should be confirmed by all positive instances of it. So a basic rule would be:

> (IC) A hypothesis is of the form "All As are Bs" is confirmed by any positive instance "Aa &Ba."[4]

A second principle is that logical equivalence preserves confirmation, which is recommended if logical equivalence, as mutual implication, is sameness of content:

(EQ) If H and H′ are logically equivalent, then if e confirms H, e confirms H′.

(For development and refinement, the classic is Hempel 1965)

However, a problem was quickly recognized, referred to as the "raven paradox." The hypothesis

All ravens are black

is logically equivalent to

All nonblack things are nonravens.

According to (IC), the latter hypothesis is confirmed by positive instances, which include blue shoes, a red jacket, yellow baskets, and the noncolorable number 33. So by (EQ) these each confirm H. The result is supposed to be paradoxical because IC and EQ are individually credible, yet they allow that finding a red jacket confirms that all ravens are black.

But is it a real paradox? The basic notion – confirmation itself – is not quantitative. Perhaps, the blue shoe does provide confirmation, only very little. Hempel also observed that simply to reject the EQ is questionable because hypotheses are involved in deductive arguments (for explanations and predictions), and validity does not vary with logically equivalent sentences.

The raven paradox has structural analogies with the Wason selection task – with "vowel" for "raven"; "consonant" for "black"; "even" for "nonraven"; and "odd" for "nonblack." The well-known findings are that to test the conditional "if there is a vowel on one side then there is an even number on the other" with cards A, D, 4, 7, subjects turn over the A card, but not the 7, and many the 4. The results, originally explained as a result of a bias in search for confirming (positive) to falsifying instances, provides insight into our response to the raven paradox: we think of looking for ravens to observe whether they are black, but not of looking for nonblack things to test for whether they are nonravens.

However, the analogy breaks down at two junctures: First, our response is presumably influenced by the background information of both the relative frequency of the different classes and of the underlying explanatory connections (a genetic account of the original hypothesis rings true, but not for the contrapositive one).

Second, in the Wason task, the pairs selected are the unusual ones for which the contradictory is as natural as the original – even-odd; vowel-consonant. But objects that are nonblack are not merely objects with colors other than black, but all manner of objects, like numbers, that could not have a color at all. The Wason task is the specialized one in which the contradictory of a natural category or predicate is itself natural.

Quine (1969) picked up on the unnaturalness of the negation of a natural class to propose that even if the two hypotheses are logically equivalent only the one with the predicate for the natural class is fit for confirmation or projection (to new cases). The notion of projectibility he drew from the far reaching ("grue") paradox of Goodman (1983). Goodman's target was a qualitative account of confirmation that attempts to model confirmation (or inductive) arguments on deductive arguments in which validity turns on their form, not their content.

The hypothesis that Goodman selected to formulate his 'grue' paradox is:

(Green) All emeralds are green.

He then offered a competing hypothesis:

(Grue) All emeralds are grue.

Something is grue, he proposed, if and only if either it is green and examined before 2050 or blue and examined after 2050. Now take any collection of positive instances of (Green) – for example, the finding of ten green emeralds each confirms, according to IC, (Green). But this same evidence will also be positive instances of (Grue), and so, by IC, confirm it. But this is the ruin of confirmation theory because (Green) and (Grue) compete (they disagree over the color of emeralds discovered after 2050), yet "grue" - type hypotheses can be manufactured for any hypothesis.

The analogy with the "curve fitters" problem is often made. Through any finite set of data points, an infinite number of functions can yield those data points. Which one should be preferred and why? If one of them is near enough a straight line, the others will appear as much more complex, if not bizarre. Still, does this basis for preference correspond to an objective feature of the world or of what we find simple?

Goodman's Paradox harmonizes with three other important theses about inference and reasoning, noted already: underdetermination, the Quine–Duhem thesis, and holism. Each of these theses, like Goodman's Paradox, are extensions of Hume's insight about induction that the evidence for a hypothesis, however strong, never implies that hypothesis. This much underdetermination is unremarkable. A slightly more exciting thesis is that for any positive evidence for a hypothesis, there are always other competing hypotheses compatible with that evidence. This is only slightly more exciting, because further evidence can be gathered, or tests constructed, to select among the hypotheses at the previous stage. A much stronger and very controversial thesis of underdetermination is that there are sets of hypotheses that are empirically equivalent (sharing the same observational consequences), so that despite their competing or conflicting, no evidence can select between them. Underdetermination has been urged against realist inferences from the confirmation of a scientific theory to either its truth or to the reference of the theoretical terms in it.

These neighboring problems lead to the question of what constraints can be applied to hypotheses so that evidence is a reliable basis to discriminate among hypotheses. In the case of Goodman's Paradox, a natural thought about "grue" is that it is artificially complex or contrived or that it predicts an arbitrary and radical change in color. Goodman anticipates this worry. He introduces another predicate "bleen": something is bleen if it is blue and examined before 2050 or green and examined after 2050. For a language with "grue" and "bleen" as primitive or simple as "green" and "blue" in our language, the latter appear complex "x is green if and only if x is grue and examined before 2050 or x is bleen and examined after 2050." From the point of view of that language, our green hypothesis is complex and it posits a radical change from objects that were grue up to 2050 to their becoming bleen after.

Goodman proposes that "green" is preferable to "grue" because it is projectible, and it is projectible because, in fact, it has been successfully used (projected) in our predictive practices. The solution, as resting projectibility on our successful habits of projection, has struck many as turning on too superficial or accidental a feature. But doubts about Goodman's solution are not doubts about his paradox as arguing for the need for prior restrictions on the predicates suitable for inductive inference and the confirmation of hypotheses, as well as the formulation of laws of nature.

11. Quantitative Induction and Confirmation

A dominant response to both paradoxes is to give up on qualitative confirmation for a quantitative approach through probability. The basic thought is that probability provides a measure of the degree of inductive support that evidence for a hypothesis transfers to it: If

$$pr(h \mid e\&b) > pr(h \mid b),$$

where b is the background or previous evidence and e is new evidence, then e is positive evidence or confirms h, and the difference between the two values is the degree of support or confirmation. (For alternative measures of degree of confirmation, see Fitelson 1999, 2003; Kyburg and Teng 2001; Tentori, Crupi, Bonini, and Osherson 2007.)

Because there are hugely many nonblack things, as well as nonravens, especially compared to the class of ravens and the class of black things, the antecedent probability of finding a nonraven among nonblack things is extremely high. Consequently, finding a nonblack nonraven little increases the probability that "All ravens are black." Understanding confirmation as above, the difference will be very small, the posterior probability becomes only slightly greater. In application to the "grue" paradox, since the prior probability assigned to "All emeralds are green" is significantly higher than to "All emeralds are grue," the former hypothesis is claimed to be much better supported by a collection of green emeralds than the latter by the same evidence described as grue emeralds. (Is this how the grue-bleen language speakers would describe the difference?)

Goodman originally directed his paradox against the earliest formulations of quantitative confirmation theory in the pioneering work of Carnap (1947; 1962) and Hempel (1965). Carnap began with the specification of a formal language over which a confirmation function was to be defined. Because there were no restrictions beyond coherence on the predicates, Carnap's formal language was ripe for the "grue" paradox, which demands such prior restrictions.

Carnap's simplest languages involved one-place predicates and names, and then a state-description is defined as the maximally consistent combination (via negation and conjunction) of all predicates and names.[5] Each state-description described a complete possible world. How should an initial probability distribution be set up among these possible worlds? Because there is no basis for taking one world as more likely than another before obtaining evidence, Carnap invoked something like a *principle of indifference* to justify assigning equal probabilities to each state-description. The arresting consequence was that learning from experience – a change in probability values with new evidence – was impossible. Consequently, Carnap proposed a set of inductive method determined by the continuum of ways to reject the assignment of equal weights to each state-description. Of these, one has the philosophical advantage that it is indifferent to bare differences between individuals, but not to kinds or properties. (See Kyburg and Teng 2001; for methods to assess these different inductive strategie's see Lewis 1971.) Carnap's plausible line of reasoning from the assignment of equal weights to state-descriptions converged on the implication of Hume's and Goodman's analyses that without prior favoring or biasing toward some "worlds" (state-descriptions) over others induction could not take place.

Conditionalization is the key principle for changing degrees of belief or learning from experience for Bayesians. Given a body of new evidence e, you should adjust your degree of belief in a hypothesis h so that it is now equal to what was (before learning the new evidence) the conditional probability of the hypothesis given the evidence. Conditionalization stipulates that

$$pr_{new}(h) = pr_{old}(h \mid e).$$

But how should you compute $pr_{old}(h \mid e)$? Bayesians take their name from Bayes's Theorem, an elementary consequence of the probability calculus. Bayesians claim that Bayes's Theorem and its role in conditionalization captures all or just about all that one requires for inductive or statistical inference, including that probabilities can be assigned to all statements. Bayesians have problems with assigning prior probabilities – what is the probability of Newton's second law before any evidence is in? If the probability is to be assigned objectively, it will have to rely on a principle of indifference. That principle runs into difficulties of conflicting representations of the set of statements to which one's ignorance is to translate into an equal distribution of probabilities. If wholly subjective, why should the assignments be accorded any respect (Sober 2005)?

In one form (where h and –h are the only two hypotheses) Bayes' Theorem states that:

$$pr(h \mid e) = \frac{pr(e \mid h)pr(h)}{pr(e \mid h)pr(h) + pr(e \mid -h)pr(-h)}$$

(Here as elsewhere, it is assumed that denominators do not equal 0.) Where there are n mutually exclusive and exhaustive hypotheses (h_i), the general formula is:

$$pr(h_j \mid e_j) = \frac{pr(e_j \mid h_j)pr(h_j)}{\sum_{i=1}^{n} pr(e_i \mid h_i)pr(h_i)}$$

The main components of Bayes's Thorem are the likelihood pr(e | h) and the prior probability pr(h). To know how likely it is that the evidence arose from h requires knowing not only of the connection between e and h, but also how antecedently likely it is that h is true. It is highly likely that if there is a snow storm in Miami that school will be canceled. Yet, the probability that the school was actually canceled because of a snow storm is still quite low. To derive the probability of h on e also requires evaluating the probability that h actually arose from e compared to the other possible hypotheses that might account for e (the denominator above).

Bayes's Theorem provides good explanations for intuitive judgments about induction such as the value of evidence that is surprising or diverse (Bovens and Hartmann 2003: Ch. 4). The posterior conditional probability of h on e, pr(h | e), expresses how much more expected h is were e true. Assume that finding e is very unexpected (its prior probability is low) (e.g. Alfred E. Neumann receives an A in his logic final). However, were h true (e.g., he had a copy of the exam earlier), e would not be surprising. So pr(AEN gets an A | AEN has the exam in advance) is high. In that case, e greatly confirms h [i.e. pr(AEN has the exam | AEN gets an A) is also high]. (This pattern of reasoning is common in inference to the best explanation discussed in the next section.)

When we move from Bayes's Theorem to Bayesianism, which restricts probability to degrees of belief or subjective probabilities, we move from the taken-for-granted to the controversial. A major advantage claimed for Bayesianism is theoretical minimalism: No claims or assumptions are made as to the basis for the initial assignment of probabilities beyond what follows from one's degrees of belief (with prior probabilities, before any learning, problematically dependent on a principle of indifference, as noted earlier). The only demand on those degrees of belief is consistency or coherence.

The minimalism is one source of criticism: any assignment of probabilities is as good as any other provided only that they are coherent (i.e., conform to the axioms of probability). Conditionalization provides for the response that if individuals assign similar values to the likelihood and conditionalize, then with new information there will be convergence on assignments of probability. But only if there is uniformity in how to judge the new information as evidence. (Another problem is how to treat old evidence newly discovered to be implied by a long-standing hypothesis or theory, so that there is no change in the probability of the evidence; Glymour 1980.)

Bayesians tend to treat all learning from experience as changes in probability by conditionalization. The treatment is in opposition to construing theoretical reasoning as aiming at the all-out acceptance (or rejection) of a hypothesis as true, detaching it from its evidential base. The Bayesians worry that the detachment is in conflict with a recognition of our fallibility, discussed later, and related problems like the dogmatism paradox mentioned in Section 5 (Levi 1980; Jeffrey 1983; Kaplan 1996).

The possibility of detachment is one way to understand a difference between probability and inductive inference. Think of a simple inductive inference, for example, Alice is an academic whose library is filled with books on early American history, so she is [probably] a historian. Strong detachment marks the end of an inquiry. The judgment or hypothesis is separated from its evidential base with all-out acceptance. Detachment implies a claim that the evidence is sufficient or representative enough for drawing a conclusion, rather than just assigning a probability on that evidence. Detachment is a risk, requiring an underlying pattern or law or regularity that is indicated by the evidence. But without taking that risk, an inquiry would never terminate, so long as the evidence leaves open the possibility of error (i.e., does not establish the hypothesis as certainly correct). The underlying regularity allows for the detached judgment, since it allows us to project from, but to go well beyond, the evidence.

By contrast, consider a fair lottery with one thousand tickets of which you own one. The probability of your losing is very high: .999. It

is reasonable to assume that this probability of your losing is higher than the probability that Alice is a historian, based on your knowledge and observation (maybe the library is simply an inheritance from her parents who bequeathed her the house?). Yet you cannot detach – all-out believe – that you will lose the lottery. Otherwise, purchasing a ticket would be worse than merely incurring a high negative expected utility. It would be to throw away money, since there is a cost without prospect of gain. Still, you can all-out believe or come to know that Alice is a historian. How come? The answer resides in the assumed underlying regularity or law in the historian case, but not for the lottery. Only the former then permits projection beyond the evidence.

An alternative explanation is that in the former case (that Alice is a historian) there is an implicit deductive argument, yielding a probability of 1, but not in the latter (that you will lose the lottery). However, even if the generalizations, regularities, and other assumptions, which underlie the inference, would, if stated explicitly, yield a deductive argument, it does not follow that the original argument is only an enthymeme for (i.e., an abbreviated version of) that deductive argument. The presumed generalization includes that conditions are normal and that in such conditions, if one is an academic and one's library is dominated by books in a certain field, then the field is one's profession. If the normalcy condition is understood to exclude all ways that the conclusion could be false, not already excluded by the explicit premise, we have a formula to convert any inductive argument into a deductive one (see Smith, Shafir, and Osherson 1993). But this does not explain the difference between the historian and the lottery arguments, any more than if we add a conditional whose antecedent is the conjunction of premises and the consequent is the conclusion, as a way to turn any argument into a modus ponens deductive one. For these additional assumptions, which are unnecessary for the high probability of the lottery cases, are themselves not backed by any grounds additional to the premises. The risk – that is, the inductive gap between premises and conclusion – is the same as the gap between the premises and these assumptions. Nothing is advanced in the argument, but only in the recognition of its basis.

If there is all-out acceptance, how should it be based on probability? A simple proposal for relating probability to acceptance (one that we have just implicitly rejected) is

(ACC) Acc(p) if and only if pr(p | e) > r, r is some suitably large value – certainly greater than .5.

But for acceptance (or detachment) more is needed. The evidence e must constitute representative evidence. Its weight of evidence – the proportion of the evidence – must be high or maximal (Keynes 1952, Cohen 1977).

Regardless of the value selected for r, Kyburg (1961) showed that a paradox follows if a conjunction principle holds:

(CR) $Acc(q_1), \ldots, Acc(q_n) ==> Acc(q_1 \& \ldots \& q_n)$.

To make it clear that the "Lottery Paradox" is indifferent to values of r (less than 1), let's set the criterion very high, say, .95. Kyburg asks us to imagine that there is a fair lottery with one thousand tickets. You have one ticket t_{123}. The probability that you will lose the lottery with that ticket is .999, which exceeds the criterion. So, in accord with (ACC), you accept the statement that you will lose the lottery. But now for each of the other 999 tickets, the probability that each of them will lose is also .999. So each of these satisfies the condition for acceptance. Because, by (ACC) you accept, for each of the one thousand tickets, the statement that it will lose, (CR) tells us that the conjunction of those one thousand statements is likewise accepted. It follows that you have accepted that all the tickets will lose. But it is impossible for every ticket to lose, if it is a fair lottery.

Both (ACC) and (CR) have come in for rejection as a basis to resolve the lottery paradox. (CR) is disputed (and here is where the Bayesian attacks) because when probabilities are multiplied (in non-extreme cases), the result is necessarily a diminution in probability. (The probability of $q_1 \& \ldots \& q_{1000}$ will typically be far below the acceptance threshold, r.) However, if acceptance is all-out, the truth of a conjunction is logically equivalent to the truth of its individual conjuncts, and so (CR) seems another expression of the basic idea of acceptance. This defense of (CR) goes along with a challenge to (ACC): probability alone is never sufficient for acceptance. As noted earlier, in distinguishing the assignment of probability from genuine inference (and detachment), no one can all-out believe that they will lose a lottery with a single ticket in contrast to the more precise assignment

of a high degree of belief. The contrast between the lottery case and genuine inductive inferences provides for an answer to the lottery paradox that is different from the one Bayesians offer, when they allow for acceptance at all (Kaplan 1996). What more than high probability (and high weight of evidence) is required for acceptance is a lingering question.

12. Inference to the Best Explanation

A natural form of inference is to infer the hypothesis that best explains the data. Sherlock Holmes infers that the owner of the dog committed the crime because this is the best explanation of why the dog did not bark when the crime was committed nearby. The data of the dog's not barking confirms the hypothesis that the owner committed the crime, and the hypothesis is introduced to explain why the dog did not bark. Inference and explanation implicate each other. Inference to the best explanation is at work where observable events, like John's rude behavior at a party, are explained by the hypothesis of an unobserved (and sometimes unobservable) cause (e.g., John was just fired from his job). (For criticisms, see van Fraassen 1980.)

Inference to the best explanation offers a promising approach to the confirmation paradoxes. The best explanation of why nonblack things are nonravens will *not* be that all ravens are black, or even that all nonblack things are nonravens. The finding of nonravens among nonblack things is best explained by the fact that nonblack things are just such a numerous and heterogeneous collection that it is unsurprising that we should find some, and so it calls for no further explanation. Similarly, that the observed emeralds have been green is well explained if all emeralds are green, which, as well, give us reason to expect the next observed emerald to be green. But the grueness of the observed emerald is not well explained by the hypothesis that all emeralds are grue, because that implies that unobserved emeralds will be blue. (For advances along these lines, see Jackson 1975; White 2005.)

Inference to the best explanation captures Peirce's (1931) "abduction," whereby plausibility accrues to a hypothesis – it is worth taking seriously. When you meet someone at a party who dresses in a similar slightly off-beat way of another acquaintance, you might think that they share political or cultural views. Although the off-beat dress is a reason to take seriously the hypothesis that the two share political or cultural views, it is hardly much of a reason to actually believe it. Suggestive or analogical reasons are important in view of the underdetermination problem of the unlimited number of hypotheses to fit any data, as well as the problem of coming up with credible hypotheses to explain complex data or observations.

Inference to the best explanation serves a stronger role as a form of eliminative inductive, whereby evidence is sought to select among rivals:

> one infers, from the premise that a given hypothesis would provide a "better" explanation for the evidence than would any other hypothesis, to the conclusion that the given hypothesis is true. (Harman 1965)

Inference to the best explanation is closest to causal explanation, where competitors are specified by contrasts – why p rather than q. When the bank robber Willie Sutton was asked why he robbed banks, he responded "Because that's where the money is." He construed the question as something such as "Why do you rob banks, rather than grocery stores?" Although having lots of money in an accessible location is only one cause for his robbing banks, it is the one that makes the distinctive difference between Sutton's robbing the bank and his robbing the grocery store. However, his questioners actually had in mind a different question, calling for an answer selected from a different contrast class ("rather than"): "Why do you rob banks rather than working a decent job?" (Garfinkel 1981).

If the best explanation is accepted as correct or even as worthy of investigation, it must at least provide a good explanation, not merely the best of a lousy lot. Criteria for the best explanation include that the hypothesis is simple, conservative, and unifying. However, the criteria should not include an explicitly truth relevant property such as high probability. That is the conclusion to be reached not assumed. The hypothesis that yields the most understanding must prove itself to be the likeliest (Lipton 2004).

Inference to the best explanation is at the center of the "miracle argument" from the success of science to a realist view of science, which hold that scientific theories are at least approximately true and that its theoretical terms genuinely refer, rather than merely serving as useful posits. The question is whether these realist claims are the best explanation for the success

of science, whose success would otherwise be a "miracle" (a grand coincidence).

13. Foundational Justification: Coherence and Reflective Equilibrium

How at the most basic level should the putative rules or principles of logic (inductive, deductive, practical) be justified?

Earlier, we noted that Bayesianism's admirable minimalism invites the criticism that it allows one to believe almost anything, so long as one conforms to the axioms of probability. It is evident that logical consistency is much too weak a criterion for rational belief – the belief that Santa Claus brings presents or that there are an even number of stars are both logically consistent. But, recall, Bayesians have a distinctive consistency requirement – coherence – which they offer to justify the probability axioms as representations of degrees of belief and preferences, and this notion they think is sufficient (with conditionalization) to represent rational belief.

The coherence arguments for commitment to the probability calculus are known as "Dutch Book Arguments." These arguments claim that if your preferences or degrees of belief do not conform to the probability calculus then a set of bets (a "book") can be designed that you are committed to regarding individually as fair, because fitting your preferences. (For references and presentation, see Hacking 2001; van Fraassen, this volume.) The bets, however, guarantee that you cannot win or that you must lose. (Conversely, if your degrees of belief do conform to the probability axioms, no "book" can be made against you.) The argument requires an idealization away from influences on betting besides an interest in winning, like the pleasure of gambling. A more problematic assumption is that logical omniscience (deductive closure) is needed, because the probability of deductive truths must be assigned the value 1.

The "sure loss" or "Dutch Book" argument can be extended to justify a principle for changing one's degree of belief by conditionalization (see Section 11). Conditionalization, recall, assigns a new probability when evidence is obtained as the old conditional probability on that evidence:

$$pr_{new}(h) = pr_{old}(h \mid e).$$

(Teller 1973; Lewis 1999; for a more general form of conditionalization, see Jeffrey 1983).

If coherence is an extension of consistency, it provides a model for the justification of the axioms of probability as representing degrees of belief that is *a priori* (not dependent on experience or sense perception). It remains an open question as to how successful are the Dutch Book Arguments and how general a model for justifying principles they provide. They do not pretend to answer Hume's problem, but they do claim to mitigate it. A principle for learning from experience is justifiable as a way to protect against a self-defeating set of preferences or degrees of belief, even if it cannot guarantee success.

Can this model be generalized to justify rules of logic (inductive, deductive, practical)? A familiar problem is that consistency or coherence is too weak, since plausible rules that are incompatible with one another can form internally consistent systems. Once substantive principles are offered besides consistency, the evident problem is a version of Hume's circularity problem (see Section 9): To establish the rules of logic requires reasoning, and how can one reason without logic (Boghossian 2000)? However, certain kinds of circularity are not a failing or not a decisive one. Proofs of the soundness of a logical principle, like the rule of conjunction simplification ("p&q" implies "p"), use an analogue of that principle in the proof. But the analogue – the soundness proof – is not syntactic or at the object level, but at the meta-level or the semantics (Dummett 1978).

A traditional view is that logic is established *a priori*. We have already noted that the Dutch Book Arguments aspire to an a priori justification for inductive principles. But Dutch Book Arguments do not reach to the foundations, since they presuppose a particular view of consistency. For logic, one approach is to claim that from reflection on, or analysis of, concepts such as "if, then," we can validate laws like modus ponens. Reflection on the concept of negation is supposed to show that instances of the $p\&{\sim}p$ must be false. This view is an advance on the one that claims that we have an faculty of intuition to discern logical laws, which merely pushes the problem back to validating the faculty of intuition. A priori knowledge is knowledge that is not based on experience, and since on traditional views they are necessary truths, such knowledge is indefeasible (no empirical findings could count as evidence against a genuine logical law).

Strict empiricists – stricter even than Hume – are not satisfied with the claimed power of reflection. They attempt a straightforward a posteriori or inductive approach to logical laws as

generalizations from observations of uniformly successful inferences. These laws are subject to hypothesis testing akin to any empirical claim (Mill 1963).

The strict empiricist approach is unpopular on Kantian grounds that observations or related evidential considerations can never establish the necessity of logical laws. An improvement over empiricism on this count is that with a logical law we cannot conceive of its failing. But what is the relation between conceivability or imaginability and real possibility (Yablo 1993)? As Putnam (1975) noted, our ability to conceive that water is not H_2O does not entail that it is really possible that water is not H_2O.

A more recent approach is to seek reflective equilibrium between principles (or rules) and judgments (or intuitions). Earlier, we noted how a soundness proof justifies a particular rule by showing that it generates only valid arguments. The semantic version of the rule is legitimate if its use is only to explain why the rule is endorsed. But for ultimate justification of a rule or principle, this explanation, which takes a lot of standard semantics for granted, is not sufficient. The justification of any set of principles eventually confronts two horns of an ancient dilemma. Either the justification process relies on further principles to justify others, which must finally be exhausted, and so return – circularly – to principles already used or, and this is the other horn, the process never exhausts itself. The process never stops. Each new principles requires a further and different principle to justify it in turn. There is an infinite regress, because these different principles must themselves be justified. The rules for, say, the derivation of excluded middle – that is, $p \vee \sim p$ – are justified by the truth-table for "$\sim$" and "v." But what justifies the truth-table (as corresponding to "not" and "or")?

Reflective equilibrium answers the latter horn by claiming that there is a starting point, rather than a regress, in the judgments of validity we already accept. It answers the former, circularity, horn by conceiving some circularities as an acceptable matter of accommodation or balance. In the richly suggestive passage that launched the reflective equilibrium model, Goodman explicitly compared the justification of inductive and deductive rules:

> How do we justify a *de*duction? Plainly, by showing that it conforms to the general rules of deductive inference. . . . Analogously, the basic task in justifying an inductive inference is to show that it conforms to the general rules of *in*duction.
>
> . . . Principles of deductive inference are justified by their conformity with accepted deductive practice. . . . Justification of general rules thus derives from judgments rejecting or accepting particular deductive inferences.
>
> This looks flagrantly circular. . . . But this circle is a virtuous one. The point is that rules and particular inferences alike are justified by being brought into agreement with each other. *A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend.* The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either. (Goodman 1965: 63–64)

Reflective equilibrium is intended as an alternative to a purely a priori justification of logical principles. Reflective equilibrium embraces neither a crude inductive justification (as just a generalization from successful instances) nor a "foundational" one, which treats some principles as sacrosanct.

In this respect, reflective equilibrium is closely aligned with coherentism in epistemology, the view that justification is holistic (Elgin 1996): A principle is justified to the extent that it coheres within one's corpus of beliefs. However, if the coherence is only within one's current corpus of beliefs, it is a narrow reflective equilibrium. It lacks the justificatory force of a wide reflective equilibrium, where one's judgments (intuitions) and principles are evaluated against alternatives principles and they are subjected to extended critical analysis. Although the approach was originally proposed as a guide to generating inductive rules, it has been invoked predominantly as the underlying model for theory construction in moral or political philosophy (Rawls 1971).

As a description of our actual practices in proposing or justifying rules, reflective equilibrium is on target. When it comes to the conditional, for instance, modus ponens and falsity when antecedent is true and consequent false are strong intuitions, with contraposition much less secure, and modus tollens in between. However, reflective equilibrium can provide little further substantive guidance. It does not specify how to balance among principles or judgments when they conflict. So it cannot settle serious disputes.

Recall that Goodman himself rejected an intuition about his "grue" paradox that most others took as a pillar.

What counts as an intuition anyway? Most of us deny the validity of disjunctive weakening (i.e., "p" implies "p v q") on initial presentation. But students are brought around by the truth-table analysis. Is the former the intuition or the latter, and how do we characterize the difference in a general way? (Is it to be assimilated to the difference between narrow and wide reflective equilibrium?) Systematization pressures intuitions. Given the interconnections that any logical system quickly establishes, principles and rules are not going to be readily adoptable or modifiable piecemeal to capture particular judgments or intuitions. Thus, to use the unintuitive rule of disjunctive weakening again, it follows almost immediately if the rules either of reductio ad absurdum or conditional proof are available.[6] Without further modification, to reject the former requires rejecting the latter. (Intuitionists and their sympathizers, however, treat the additional losses as a bargain.)

14. Paradoxes

Reflective equilibrium is a practical approach to the evaluation of reasoning principles, offering no latitude for sceptical doubts about the ultimate justification of our principles (or intuitions). A search for reflective equilibrium is impatient with paradoxes, like the sorites, because one option in the face of paradoxes is a sceptical or nihilist conclusion that certain basic concepts are ultimately incoherent, even if we seem to make intuitive sense in talking about them.

Paradoxes involve the derivation of either a contradictory or a starkly unacceptable conclusion, like that there are no heaps or rich people, from prima facie valid chains of reasoning, derived from apparently sound premises. A solution to a paradox would be to uncover an error either in the premises or in the reasoning. Difficulties in solving the paradoxes, particularly ancient ones like the sorites, raise sceptical prospects.

If not the sorites, then the Liar, is the paradox that has occasioned the most intense investigations. Tarski's equivalence condition earlier (see Section 2) was formulated in his response to the Liar Paradox:

(A) Sentence A is false.

From an intuitive assumption about any sentence, (A) is either true or false (or expresses a proposition that is either true or false). If it is true, what it claims is correct. But what it claims is that (A) is false, which is, effectively, "I am false." So, it's false. If, however, it is false, then what it claims is wrong. So it is true. In short: (A) is true if and only if (A) is false.

Like the sorites, the Liar Paradox encourages suspicion of bi-valence (the thought that every sentence is true or false), while supporting logics allowing for truth-value gaps, as well as gluts (more than two truth-values) However, if you think that this move alone will do the trick, out pops a "strengthened Liar"

(A′) Sentence A′ is not true.

One can claim that denying "Sentence A is false" does not necessarily imply that Sentence A is true. But this is no escape from the strengthened Liar, because even if "not true" is weaker than "false," as the strategy requires, the contradiction arises when we ask whether (A′) is true.

Focusing on a version like (A), Tarski thought the culprit is that "semantically closed languages" (i.e., those which contain their own truth predicate and a device – like naming – to denote each sentence) are inconsistent. In those languages, not every sentence can be assigned the value *true* or *false*. Part of Tarski's solution is to define truth only for open languages, establishing a hierarchy: truth for language L is defined in a higher (meta) language, which has the resources to pick out all the sentences of L. Tarski showed how to define truth for a language (in a metalanguage) by way of recursive definition. From a finite assignment of truth to atomic sentences, sentences of any complexity can be generated, for example, "A and B" is true if and only if "A" is true and "B" is true. (For critical discussion Field 1972; for applications to natural language Davidson 1984. Kripke (1975) contains both an example of consistent predications of truth that receive contradictory assignments in Tarski's hierarchy and an alternative way to define truth.)

One problem with Tarski's approach to our ordinary notion of truth is that we can understand sentences that predicate truth without any knowledge of their truth-level in his hierarchy. A suggestion here is that "true" is implicitly indexical, like "I" or "you," whose reference is determined within the context of utterance, although the meaning is constant (Kaplan 1989). Thus:

(A) Sentence A is not true$_1$

and

(A) Sentence A is $\text{true}_{|+1}$

are not contradictory, anymore than "You are a lefty" and "You are a righty" affirmed of different people (Burge 1979).

A paradox of set theory, whose origin lies with Liar-like self-reference, was discovered by Russell – hence, "Russell's Paradox." Typically, members of a class, like members of the class of people, are not themselves classes and so they cannot be members of themselves. But some classes are: The class containing as members all classes that have more than five members. Russell asked a question that combined these: Is the class of all classes that are not members of themselves a member of itself? If so, then it meets the defining conditions that it is not a member of itself, so it is not a member of itself. But if it is not a member of itself, then it satisfies the condition, and so must be included as a member of the class of all classes that are not members of itself. In brief, the class of all classes that are members of themselves is a member of itself if and only if it isn't.

Even though Russell's Paradox is one of set-theory and the Liar one of semantics, they have striking similarilities, which led Russell to compare them. The contradiction yielded by either involves similar reasoning from respective self-referential questions. Both involve extremely plausible existence claims about the universal predication of their respective key notions (truth-value; membership) and invite similar hierarchical restrictions on that predication. The problem is not self-reference alone. Numerous self-referential statements are unproblematically true ("This sentence is in English") and others unproblematically false ("This sentence is in Russian"). Russell thought that the essential connection was, roughly, that a member of a totality is defined in terms of the totality (the Vicious Circle Principle).

A number of the central paradoxes share a feature close to the fault line that Russell proposed. These paradoxes claim the existence of a meaningful property or term to fit a specified general condition. The sorites requires that to every concept there is a determinate answer to whether the concept applies to an object or not. The very different "grue" paradox still has this feature: It casts doubt on the condition that to every predicate there corresponds a confirmable or projectible property. The Liar Paradox assumes the condition that to every grammatically well-formed sentence there is one of two truth values. Russell's Paradox is sometimes illustrated by the Barber Paradox. In a town, where the barber shaves all and only those citizens who do not shave themselves, who shaves the barber? A Russell's Paradox-like contradiction follows. But the Barber is only a contradiction, no paradox. For no credible principle establishes that there is such a barber. Not so for Russell's Paradox. Russell originally directed his paradox to an explicit axiom of Frege's system, which implied the assumption tacitly made in constructing the paradox: To every coherent condition there is a set meeting exactly that condition.

The natural reaction to Russell's Paradox is to deny that the membership in the set of all sets that are not members of themselves is a coherent condition. But the denial does not sit well with much positive work in set theory and, specifically, if the restriction suggested is not qualified it undermines other fundamental results (e.g., Cantor's diagonal proof that the power set of a class has more members than the class, a proof that shares structural features with many paradoxes, including the Liar and Russell's).

15. The Preface Paradox, the First Person, and Fallibility

The Preface Paradox is a paradox of self-reflection that arises naturally from thoughts about one's fallibility. Its name derives from a typical disclaimer that occurs in prefaces to books (Makinson 1964). The author of a non-fictional work writes something intended to express modesty "Remaining errors in the book are my own." The author is justified in making this assertion. Given the numerous statements in the book, he is bound to have made some error. Assuming the author is conscientious, however, he believes each sentence he wrote and asserts them to be true. But it is impossible for each statement in the book to be true, if the preface is part of the book.

The bite of the Preface Paradox arises when we realize that fallibility about one's own beliefs is akin to the modesty expressed in the preface. In taking oneself to be fallible one appears to believe that "at least one of my beliefs is false." However, for each of one's beliefs, one does (trivially) believe it. But one's corpus of beliefs, including the belief in one's fallibility so expressed, is inconsistent. As standardly presented, the Preface Paradox has the form of "w-inconsistency." All instances of a generalization are true, but not the generalization itself.

Unlike other paradoxes, the dominant view is that the consequence of the Preface Paradox is acceptable. In maintaining this corpus, a fallible believer is "being rational though inconsistent" (Makinson 1964: 207).

But can we just resign ourselves to this consequence? The result would be a kind of complex Moore's Paradox: "p_1, $p_2, \ldots p_n$, [are each true] but I believe that not all of p_1, $p_2, \ldots p_n$[are true]," where among the p_i is the quoted sentence itself. A different way out is that even those who admit their fallibility do not really all-out believe that at least one their beliefs is false. Rather, they take it to be only extremely probable in parallel to the impossibility of all-out accepting that one will lose a fair lottery with a single ticket.

How should one factor into any of one's reasoning (inclusive of that reasoning) one's belief that one is fallible? Hume (1978) observed that if we incorporate into our judgments an evaluation of those judgments that represents our fallibility, a sceptical regress threatens. Hume writes:

> In every judgment, which we can form concerning probability, as well as concerning knowledge, we ought always to correct the first judgment, deriv'd from the nature of the object, by another judgment, deriv'd from the nature of the understanding. . . . In the man of the best sense and longest experience, this authority is never entire; since even such-a-one must be conscious of many errors in the past, and must still dread the like for the future. Here then rises a new species of probability to correct and regulate the first . . .
>
> Having thus found in every probability, beside the uncertainty inherent in the subject, a new uncertainty deriv'd from the weakness of that faculty, which judges, and having adjusted these two together, we are oblig'd by our reason to add a new doubt deriv'd from the possibility of error in the estimation we make of the truth and fidelity of our faculties. . . . But this decision, tho' it shou'd be favourable to our preceeding judgment, being founded only on probability, must weaken still further our first evidence, and must itself be weaken'd by a fourth doubt of the same kind, and so on *in infinitum*; till at last there remain nothing of the original probability. (182–183)

This argument is puzzling even aside from its skeptical end, because it seems only to require a recognition of one's fallibility and the willingness to take that into account to qualify one's judgment. However, once a judgment is made, it is made, and to do it over in light of one's fallibility is to just double-count. Moreover, why should the correction – representing one's fallibility – lessen the original probability, rather than raise it? After all, fallibility is about error, and one can err in either of two directions – overestimate, as well as underestimate. Fallibility bears on the weight of evidence, not its force. (In familiar psychological studies, subjects are asked, e.g., "Which is more populous city San Diego or Santa Fe?" After providing that answer, subjects are asked, "How confident are you in your prior judgment?" Hume's conflated question now is evident: "How probable do you now think it is that Santa Fe is larger than San Diego in light of your prior judgment of fairly high confidence in your even earlier judgment than Santa Fe is larger than San Diego?" [huh?].) Consequently, it does not make sense to calculate the probability of a judgment by an integration (via Bayes's Theorem) with the weight of evidence, where this is measured in Hume's terms, by one's estimate of one's degree of fallibility for the judgment at hand.

16. Fallacy and Charity

If the objections just given succeed, Hume's argument is fallacious. A fallacy is an argument that seems to be good, but which isn't (Hamblin 1970). One is falsely persuaded, unlike the case of crudely bad reasoning (e.g., If John is in Albany, he is in New York. So even if he isn't in Albany, he is in New York.).

The following is an example of a student committing a fallacy of denying the antecedent (i.e., "If p then q; not p; therefore, not q"):

(26) If I don't study, my dad won't let me go to the party.

So, (27) I'll study, then he has to let me go out.

More common and seductive are fallacies of scope:

(28) Necessarily, if Jeff is a teenager, he is under twenty.

(29) Jeff is a teenager.

So, (30) Necessarily, Jeff is under twenty [i.e., he cannot become an adult].

Or, (31) Alice does not believe that the library is open on Saturday.

So (32) Alice believes that the library is not open on Saturday.

But are these common fallacies? Would not anyone who reasons to (30), really mean that Jeff must be under twenty only given that he is a teenager? Assumptions about the necessity of human rationality have led to scepticism about whether people ever really commit fallacies or that there really are common fallacies. (A weaker reading is that the attribution of fallacies is never warranted. Problems with the former, which we concentrate on, apply to the latter.) The best defense of this conclusion is that when we understand the speech-acts of others, we must see them as rational, hence, we must interpret their assertions under a *principle of charity* (Davidson 1984). Think of the comprehension of metaphor or hyperbole or similar figurative speech, along lines already suggested by Gricean pragmatics (see Section 7). A recent newspaper column states "Bush is radioactive." If the political commentator who wrote it really believed that President Bush is subject to radioactive decay, he would have to be so bereft of rational thought that he would be unintelligible, including to himself. Instead, the striking falsity of a literal reading of that utterance is a signal to infer that the best explanation of what is meant is something different, for example, if a political leader associates closely with President Bush, the association immediately taints that leader. Now the speaker can be readily viewed as rational (Davidson 1984) and as speaking in accord with the cooperative principle (Grice 1989).

Because a fallacy is a serious failure of reasoning, the principle of charity, as well as the cp, are alleged to undermine a fallacy reading, especially, where an evident nonfallacious alternative is available. In (26)–(27), the first premise, although stated as a conditional, is reasonably meant as a biconditional and then the inference goes through.

Scepticism about the commission of fallacies is unwarranted. First, when someone commits a fallacy, the person need not, even cannot, recognize it as having a fallacious form (e.g., if A, B; so if not A, not B), let alone being a fallacy (which would be to commit a Moore's Paradox: "My argument establishes its conclusion, but the argument is fallacious"). Second, the principle of charity, assumptions of rationality, and the cooperative principle assume a background of mostly well-founded beliefs and good reasoning. One can commit many fallacies compatible with this assumption, because the "mostly" is for a much larger reference class of beliefs and reasoning than the "many" (salient or prominent arguments). The former include such banalities as that: If today is Tuesday, I have class. Today is Tuesday, so I have class. Or, Jim is not in the classroom. So, Jim is somewhere else. You can easily understand and communicate with a person, who commits the above three fallacies, because his reasoning remains predominantly cogent.

In fact, one can readily explain on rational grounds why the above and other seductive fallacies are committed. It is easy enough to construct a reasonable path (from a misunderstanding of the law of large numbers) to the committing of a gambler's fallacy. Or, take (26)–(27). The student understands the relevant words and sentences he utters, which already presupposes a huge amount of rationality. (The student realizes that his father takes seriously that he should study, that this is a domain of his father's authority, and that his going to the party depends on his father's approval.) The student also understands that many times when he has done what the father regarded as his duty, the father allowed him to engage in a desired activity only when the duty was performed. Factor in to this attempt at rational explanation for fallacious thought normal human distraction and indifference, as well as our inclinations to impose contextually invited assumptions. It is not then at all uncharitable to ascribe to the student an inference on these occasions to the belief that whenever he fulfills a necessary condition for gaining his father's permission, he has thereby fulfilled a sufficient condition, although the student probably would not describe it in these ways (of "necessary" and "sufficient" conditions). Finally, a simple test applies. Would the student who reasons from (26) to (27) be surprised and taken aback, rather than hostile to an unfair accusation, by the following argument as paralleling his own in form and yet clearly (to his mind) fallacious?:

(33) If I don't study, I won't get into Yale.

So, (34) I'll study, then I must get into Yale.

17. Implicitness and Argument

These examples all illustrate one facilitator of fallacies, compatible with our rationality and basic reasoning competence. Much of our reasoning and particularly our ordinary arguments are tacit or implicit as compared to the reconstructions of those arguments, where missing

or hidden assumptions are stated. The student would likely not endorse the abstract logical principle his argument assumes: if *p*, *q*. So, if not *p*, not *q*.

This observation provides no formula for eliminating common fallacies, because ordinary argument and reasoning cannot aspire to much explicitness. Implicitness is a barrier that can be overcome in specific cases, but it is ineliminable, for reasons of economy and limited grasp of our beliefs. Tacitness or implicitness is evident in the drawing of conversational implicatures, allowing communication of much information with brevity (see Section 7).

The extent of the implicitness even of deductive arguments is obscured when they are reconstructed so that the validity is exhibited in the logical form alone as with the opening example (1)–(3): Either *p* or *q*. Not *p*. So, *q*. There are valid arguments, whose validity does not reside in their logical form:

(35) John is taller than Jim.

So, (36) Jim is not taller than John.

The argument's validity pivots on the asymmetry of "is taller than." But asymmetry is clearly not a property of all relations or two-place predicates (cf., "as smart as"). To capture the validity of (35)–(36) by logical form, there would need to be a separate premise affirming the asymmetry of "is taller than."

In everyday reasoning rarely would such a premise be explicitly thought. If arguments or inferences are not valid by their logical form alone, then arguments such as (35)–(36) can be valid as they stand (Brandom 1994, Ch. 2; Thomson 1965). It is valid as a result of the content of its substantive or nonlogical vocabulary ("is taller than"), rather than only because of its logical form as with (1)–(3).

Accepting the validity of (35)–(36) as it stands is one way to introduce the implicit nature of much reasoning, though greatly understating it. The following would be quickly recognized as a good argument, although stating the assumptions informing it explicitly would be burdensome:

(37) North Korea successfully tested missiles today.

So, (38) the United States is going to call on the United Nations to impose sanctions.

The argument draws on the audience's vast background knowledge to efficiently – without stating – provide a bridge to the conclusion.

Although brevity and limited knowledge are fundamental reasons for an ineliminable implicitness of argument and reasoning, persuasion is aided by it. The one persuaded will supply without recognition the missing assumptions, and so he commits himself to them. The benefit to persuasion may not be intended, but simply a product of ignorance, because we often do not know what are the assumptions our arguments require. An argument that used to be found persuasive is

(39) Scientists discovered that the Morning Star and the Evening Star are Venus.

So, (40) it is a contingent matter that the Morning Star and the Evening Star are one and the same.

We now realize that the argument rests on an assumption:

> If it is an empirical discovery that *p* (i.e., if from the point of view of the investigators, it might have turned out otherwise), it is contingent that *p* (i.e., *p* could actually turn out false).

Once the assumption is stated explicitly, it immediately appears suspect. Our missing the invalidity of the inference is evidence that until Kripke's (1980) discovery we did not even recognize that an assumption was made.

A conceptual basis for implicitness – in particular, that an argument cannot express as premises all the reasons logically involved in deriving its conclusions – is the moral of Lewis Carroll's (1895) parable "Achilles and the Tortoise." Achilles wants to show the Tortoise that anyone who accepts (41) and (42) must accept (43):

(41) Things that are equal to the same are equal to each other.

(42) The two sides of this Triangle are things that are equal to the same.

(43) The two sides of this Triangle are equal to each other.

The Tortoise accepts (41) and (42), but not (43). Achilles agrees that he must now show the Tortoise that

(44) If (41) and (42), then (43)

But what if the Tortoise responds that he accepts (41), (42), and (44), yet he still does not accept (43)? Then Achilles must show him, by parity,

(45) If (41) and (42) and (44), then (43).

Clearly, Achilles has been persuaded to enter an infinite regress. The moral of the story or one moral, anyway, is that Achilles should have balked at the first step: (41) and (42) already imply (43), so that no further premise is necessary to justify that inference. For the Tortoise to claim that he accepts the conditional (41), as well as (42), but that he does not accept (43) is to contradict himself (the validity of the inference is implicit in the meaning of the conditional). He is committed to (43), even if he lacks any specific belief in modus ponens as a valid argument form.

Despite the essential benefits to reasoning of implicitness, the reconstruction of arguments in which assumptions are stated and terms and statements are standardized is central for purposes of the critical analysis of arguments. By rendering these assumptions visible and as entering claims to truth, burdens of proof are imposed on the claimant. Explicitness renders bad reasoning and confabulation more difficult.

Explicitization is a natural consequence of argumentation as a dialectical exchange. Because argumentation is social – ideally, public – to engage in it is to impose a questioner or critic or interlocutor on oneself, and so a device of self-correction. If the interlocutor is not under the arguer's control, if he does not share the same biases as the arguer, he is thereby an obstacle to bending argument to favor the arguer's own position. Socratic questioning compels focus by interlocutors on crucial consequences or implications, keeping the dialectical exchange on target. Of course, explicitization can also add biased information. It can overload and distract one's thoughts. Interlocutors may be heavily self-selected. But these are interferences with argumentation, and our purpose here is to highlight its potential value functioning optimally for reasoning (Adler 2006).

Argumentation involves moving one's argument from its inchoate form to an articulate one. For complex arguments, argumentation almost certainly calls for written formulation, which is characterized by greater explicitness, since addressed to a more impersonal audience (than in casual thought or conversation). In articulate forms, the social activity of argumentation can become public, with unrestricted access. In argumentation, participants attempt to render explicit implicit assumptions, inferences, qualifications, and so on. Explicitization brings forth the basic structures or generalizations that constitute the underlying warrants for inferences. (On warrants, see Toulmin 1958: Ch. III; this volume) It facilitates reconstruction of the argument with variables, which encourages testing in disparate domains, through diverse substitutions for the variables.

18. Social Reasoning and Oneself Over Time

Argumentation is a form of social or dialectical exchange. In one form, close to Socratic dialogue, one agent acts as questioner of another's claim or thesis and elicits his reasons for that claim or thesis. The objective is to test whether that claim or thesis and the reasons for it are consistent, including with other claims (beliefs).

Besides its value for self-correction and learning, argumentation and social reasoning, more generally, foster coordination with others or oneself in the future and they exploit a division of epistemic or cognitive labor. In argumentation, we take turns as questioners and claimants (arguers) and by presenting arguments to others we automatically draw on information and skills that they have, which we do not.

Although we cannot draw on information that we will obtain only in the future, we can alter beliefs as new information comes in, which improves the beliefs of our later self. Recall that conditionalization is a primary mechanism for updating probabilities with new information, generating a rational connection between earlier and later stages of oneself (see Section 11). An analogue of conditionalization in practical reasoning does not appear as secure. Suppose the deterrence of a nuclear war is expected to succeed if the United States provides evidence that it will retaliate. Then the deterrence effect is a reason to intend to retaliate. However, this reason, though sufficient to justify the attitude, may not be a reason sufficient to render it true that one will retaliate. Were the deterrence to fail, it might be better to enter negotiation rather than engage in all-out nuclear retaliation (Gauthier 1986; Kavka 1983). The transition from the conditional intention to the all-out intention, when the condition is realized, ought not to be taken for granted. Is conditionalization subject to anything like this gap?

The gap involves how one's present self seeks to guide itself over time. The future self subject to the conditional intention does not regard itself as bound by it. Surely, I am not bound now to agree with what I regard as my earlier foolishness. (However, the earlier judgment could be a strategic one, adopted with the intention that one's later self should not be faithful to it.

More on this strategic explanation later.) Conditionalization seems applicable. I may recognize a continuous chain via conditionalization to my present position. But I may now regard the limited information and foolishness of my earlier self as implying that my past assignment of probabilities places no restrictions on my present self. I do not conditionalize from the position of (what I now believed are) earlier immature thoughts.

However, when we turn to our understanding of ourselves in the future – our future selves – a surprising principle has, like conditionalization, also been defended by a Dutch Book (or "sure loss") Argument. This is van Fraassen's "Reflection Principle," which says, informally, that if your probability now that A at some later time is r, then right now the probability that you assign to A should be r. More formally:

$Pr_t(A \mid pr_{t+x}(A) = r) = r$ (van Fraassen 1984).

The principle implies a commitment to consistency with one's later self, not merely an expectation that one's later self will be more informed or mature. The commitment holds even if one regards one's later self as subject to prejudices or biases (e.g., one's present youthful liberal self anticipates that with middle age one will become, unfortunately, less liberal and more politically conservative), unless one ceases to identify with one's later self.

This principle seems to say that I cannot regard my future self as foolish as I can my past self. Can't I, and if not, why? (It does not seem to be a Moore's Paradox to affirm "Taxes to support welfare is just, but *I will* (in my conservative middle age) not believe it.") (Bovens 1995) Is this a problem for the Reflection Principle? Alternatively, does it show that "sure loss" arguments, which turn on preferences as fixed only by one's judgments of what is likely, are not applicable to all-out beliefs or acceptance? All-out acceptance is responsive to our need to economize – to end inquiry in a finite amount of time, despite the possibility of acquiring further evidence.

The hold of my present self on my future self is central to practical reasoning – one's present self determines how one's future self should act. Once a goal or end is fixed for practical reasoning, though, what form do the conclusions take? When a student reasons about whether she should talk to a teacher to question a low grade, her conclusion is not merely that on the evidence, she should go to talk to him. That does not yet determine how she should act – perhaps, she should find more evidence? Her conclusion is a commitment or intention to act accordingly – to speak to him (or not) – akin to all-out acceptance. Although sometimes practical reasoning does not reach a terminus and further inquiry is called for, the aim is a decision to act, when feasible.

A natural principle of rationality is that if I conclude that I ought to A (e.g., floss my teeth), all things considered, then I form the intention to A or I act to A. But this natural principle is one we often disobey. My concluding that I ought to floss tonight is a far cry from my deliberately forming the intention to floss. When I judge that I ought to floss all things considered (including my laziness and displeasure with flossing), but I do not intend to floss or I deliberately or intentionally do not floss, I suffer weakness of will or *akrasia*. Because weakness of will smacks of irrationality, perhaps the cases that seem to fit it are really ones where at the moment of action, I change my judgment to a denial that I ought to floss, rather than just relax in bed. Attributions of irrationality are hard to square with actions that appear to make sense. There is then a drive to avoid the attribution by redescribing what is going on as a change in judgment (a form of the "principle of charity"). Still, not all cases can be redescribed in this way. When I do not floss, I have not changed my mind that I am really better off not flossing. I suffer weakness of will, allowing laziness to thwart good judgment (Davidson 1980).

But there are some cases in which the conclusion of practical reasoning does not yield the corresponding intention or action, and there is no irrationality. A teenager reasons that because a teacher purposely embarrassed him in class, he should get even by puncturing the tire on the teacher's car. The student reaches this conclusion, yet when the moment of truth comes he does not act. Nevertheless, his weakness of will is not irrational, given the teenager's overall values, beliefs, and interests. His practical reasoning simply did not comprehend all his genuine reasons for actions in the circumstance, some of which provide forceful resistance to the teenager's proposed action.

What force is there then in speaking of commitments, decisions, or intentions, over and above (detached from) the relation of the evidence or reasons to the conclusions as to how one ought to act? In either case one ends inquiry with a judgment and sets up a barrier to reconsideration. When one commits oneself, however, one takes a stance that one will not act

otherwise, regardless of what one learns in the interim except in extreme cases. That stance goes beyond merely drawing a conclusion as to how one should act. A promise is the most familiar example: If I promise to meet you for lunch next Tuesday and in the interim I receive another invitation, I will not accept it. Of course, if an emergency comes up in the interim, then I do legitimately break the engagement (Raz 1990).

When promising, as when one forms an intention or commitment, one's present self generates reasons that bind one's later self from acting otherwise. Given that one's later self will acquire further information, unavailable to one's present self, including detailed information about the current circumstances, how can it be rational to so bind oneself in the future? Answers to this question, which also go a good way to answering an analogous question for why we should ever adopt rules, center on coordination with oneself and with others and anticipated limits on rational judgment. If I decide that I will be in London in June, rather than merely regarding that as my best option so far, I can form plans around that trip (e.g., to purchase tickets to a London play). More important, if I commit myself to act in certain ways, others can coordinate their plans with me. We can arrange for a meeting on the presumption that we will all attend, which we could not do if inquiry or options remain open.

In binding one's future self, one forecloses one's later self from exercising its will. Most vividly and with highly variable success, we commit ourselves to diets that deny our later self from acting deliberately on its own. How is the binding to work, because at that later time, I can simply decide to follow my current judgment, rather than my past one?

One famous model for these problems is the tale of Ulysses and the Sirens (Elster 1984). Ulysses wants to hear the song of the Sirens, but he wants to avoid the dangerous madness that hearing their song is known to induce. His solution is to have his men tie him to the mast as they pass the Sirens' island, and to stuff their ears so that they can hear neither his cries for release nor, more importantly, the Sirens' song, as he can.

However, the irreversible binding is risky. What would happen if Ulysses' sailors spy a storm or an enemy ship, while he is bound? They acquire information unavailable to Ulysses' present deciding self. They need Ulysses' guidance, but only if the sailors release him, could he provide it. If the sailors follow the earlier command, it will be to the ship's detriment. Alternatively, the new determination counts as an emergency circumstance cuing them to release Ulysses and to unstuff their ears. But this alternative reintroduces a role for his sailors' judgment and thus the original problem of how to maintain the decision of the earlier self.

Irreversible or absolute binding does not, in fact, cover many realistic cases where the binding is porous (e.g., "controlled cheating" in diets: you can occasionally eat a sweet for dessert) and we add incentives to lessen the benefits of the immediate reward (e.g., if you lose a certain amount of weight, your family takes you on a vacation) (Ainslie 1992).

Commitments that are less binding than that of Ulysses are often made to avoid temptations that one judges overall worse for oneself. You purchase a subscription to the theater, even when you believe, let us assume, that you will have no problem purchasing a ticket for performances on the selected evenings. You do so because you value going to the theater and you know that your own laziness on the night of the performance will lead you just stay home and watch TV. The subscription pressures you to go because otherwise you will have wasted money, and so you are very likely, as you anticipate, to go to far more plays with the subscription than without it.

But wait: If on a given night, you genuinely prefer to stay home, why should it make any difference whether you have a subscription or not? The money is spent in any case, and since without the subscription you would not have gone, why should you allow the past to force you to do what you currently disprefer? Economists refer to these past investments as "sunk costs," and they recommend that we do not honor them. It is irrational to do so, if our only goal is to maximize profits. In that case, future return is all.

But is maximizing profit all? We value committing ourselves to a plan and sticking to it. For others to trust us, requires that when we enter a commitment to another – say to meet them for lunch – we honor it. In this way, one gains a reputation for trustworthiness, which invites others to continue to cooperate, an invitation that is foregone if one regularly subjects one's commitments to the test of better offers and future returns (Nozick 1993).

The honoring of commitments and the maintenance of trust are necessary for stable solutions in "Prisoner's Dilemma" (PD) situations. In a standard example, we agree to a long-distance trade – your $1,000 for my stereo – and we will realize the trade by your sending a check

to me and my shipping the stereo to you. If the trade goes through, we are both better off, since I value your $1,000 more than my stereo and you value my stereo more than your $1,000.

There are, however, two contrary pulls: fear of being a sucker and the prospect of far greater gain. If you do not hold up your end of the bargain and I do, I am out a stereo without compensation; and if you hold up your end of the bargain, I realize far more if I do not send you my stereo. Now if the trade is made through any legal institution, like e-Bay, each participant has a legitimate fear of punishment if he reneges. However, the recourse to an external authority, which Hobbes took as the only way to resolve the dilemma, should be unnecessary. Isn't rationality enough to allow us to realize the mutual, if second-best, gain of cooperation?

The problem is represented vividly in the following matrix:

		You	
		Cooperate	Not Cooperate
Me	Cooperate	A. Me: + 100; You: + 100	B. Me: −100; You: + 200
	Not Cooperate	C. Me + 200; You: − 100	D. Me: 0; You: 0

Except in special cases, these numbers, which are to represent utilities, not monetary value, do not matter only the ordering:

For me: C>A>D>B
For you: B>A>D>C

Our best and worst outcomes are opposites; and our second and third best match. As a consequence there is a "Dominance" argument not to cooperate. From my point of view, C is better than A and D is better than B. Although the Dominance argument is strictly rational, its Achilles Heel is that if you reason as I do, we both wind up at D, which is worse for both than if we cooperated (A). These two opposed arguments – for and against cooperation – are what make the PD a dilemma.

The problem arises as well in many person PDs, for those who would "free ride" – gain the benefits of the cooperation of others without sharing the sacrifices. A large community is conserving water in order to avoid a drought. (Another obvious example is voting.) You know that most others will conserve, and that if most others do, there will be no need for rationing water, which will be far worse for all. Because there are so many persons who are involved, if you defect by taking an extra long shower, that will not be noticed, so you free-ride. As long as few of you defect, the community does stave off the drought. In fact, if few defect this is presumably better overall, because it is sufficient for conserving that most, not all, cooperate, even if the really cooperative result would be to rotate opportunities for longer showers. Free-riding will only succeed for you however if not only does it remain fairly secret that you defect, but it remains secret that any group or collection is free-riding. However, the pull toward free-riding is available to anyone who reasons as the defectors do. The conservation policy will then be undermined and all will be worse off.

The cooperative solution, whether in the two-person or the many-person PD, is a case where self-interest and ethics harmonize, giving a positive answer to the ancient question "Why should I be ethical?," treating cooperation as the ethical result. Cooperation benefits each cooperator. It's in their interest to do what is ethical. In a two-person PD, we are both better off cooperating than if, instead, neither of us do, both acting on behalf of immediate, rather than enlightened, self-interest.

Outside the bounds of an external authority, realizing the cooperative result depends on trusting other participants, which thereby leaves us vulnerable if another is untrustworthy or chooses to defect. In a repeated Prisoner's Dilemma, as with ongoing trades or meetings, there is much greater prospects for trust and so the cooperative option. We each know what the other did previously, giving each the power to retaliate or to defect on the next round. Each of us recognizes the benefits of maintaining our cooperative practice for which either one can opt out, as well as defect. The strategy that is overall most successful is "tit-for-tat." If you play tit-for-tat you start off by cooperating and continue to cooperate unless the other party defects. After which you defect in retaliations, but then return to cooperating, rather than holding a grudge. The strategy is simple and it is easy for others to recognize that you are playing it, which are among the main reasons for its success (Axelrod 1984).

The Prisoner's Dilemma in its one-shot form bears unexpected affinities to another dilemma that Nozick (1969) introduced: Newcomb's Problem. In Newcomb's Problem, the Superior Being, whose predictions have always come to pass over many trials, offers you a choice to be made one week later between taking the contents of Box 1, which contains 0 or $1,000,000 only or taking as well the contents of Box 2, which contains $1,000. If the Superior Being predicts that you will take the contents only of Box 1, he places $1,000,000 in it; otherwise, if he predicts that you will take the contents of both boxes, he does not put anything in Box 1, and you wind up with only $1,000. The argument to take only the contents of Box 1 is evident enough: you secure $1,000,000. The argument to take the contents of both boxes is also evident: Because the being has already placed the money in Box 1 or not, you have nothing to lose now – one week later – in taking the contents of both boxes.

In Newcomb's Problem, your action of selecting a box is causally independent of what the Superior Being does (assuming no reverse causal process). But there is a probabilistic dependence, since what you now do affects the probability of the reward that you will receive. (In the PD, the actions of each party is assumed causally and probabilistically independent of the other. Were there probabilistic dependence, the PD gives rise to a similar conflict.) Then there is a dominance argument to take both boxes:

		Superior Being	
		Predicts One	Predicts Take Both
You	Take One	$1 million	$0
	Take Two	$1 million + $1000	$1000

Taking the contents of both boxes dominates for you over just taking the contents of Box 1 (the box, which you cannot see into, that has $1 million or $0).

However, assuming that the probability of the Superior Being is very high (e.g., .95) and that your utility is roughly the same as the dollar rewards, you maximize expected utility by taking only Box 1:

$$EU(\text{Box 1 only}) = (1{,}000{,}000)(.95) + (0)(.05) = 950{,}000$$

$$EU(\text{Box 1 and 2}) = (1{,}001{,}000)(.05) + (1{,}000)(.95) = 51{,}000.$$

Another twist, which favors a two-box choice, is to limit the expected utilities to causal dependencies of the action taken and the results, which, again, is assumed not to operate backward in time. For a typical application: Assume that smoking and lung cancer have a genetic base, but that the only cause of lung cancer is genetic. Smoking and lung cancer remain probabilistically dependent. Then it would seem that one should not quit smoking to avoid lung cancer, assuming that smoking is pleasurable. Yet to do so does lower the probability that one has lung cancer (Gibbard and Harper 1978; Lewis 1986a; for a collection on the Newcomb Problem and the Prisoner's Dilemma with an excellent introduction, see Campbell and Sowden 1985; also Sainsbury 1988, this volume).

Earlier, we mentioned a strategy that you would like to adopt here. You would like to intend to take the contents of only Box 1, up to the moment when the Superior Being makes his prediction. But then, when your moment of decision comes (one week later), you actually do not follow through on that intention. You then take the contents of both boxes. You secure the prediction you seek from the Superior Being without the real commitment to it. However, this strategy confronts the obstacle that the Superior Being may see through your facade. So you have a new version of the old conflict: a dominance argument to take both boxes and an expected utility argument to take only the contents of the million dollar box. What should you do?

Notes

1 In this chapter, we use the symbols "&" (and), "v" (inclusive or), "~" (not), and "⊃" (the material conditional) was the meanings given in standard texts on classical logical logic.

2 A nice example to distinguish indicative from counterfactual (subjunctive) conditionals is from Adams 1970: on our current understanding, the indicative that if Oswald did not kill Kennedy, someone else did, is true, but the counterfactual that if Oswald hadn't killed Kennedy, someone else would have, is false.

3 On this reading, Hume's argument is a far-reaching scepticism:

> Even after the observation of the frequent or constant conjunction of objects, we have no reason to draw any inference concerning any object beyond those of which we have had experience. (1978: 139)

4 There are examples that violate IC e.g., the hypothesis is that ravens are under 5' long, and you find a raven that is 4'10". This and similar examples depend, however, on background

information, which the original presentation of confirmation theory attempts to abstract from.

5 For example, suppose that there are just two predicates in our universe, Red and Square, and just three names, Huey, Dewey, and Louie. Then a state description is an assignment of each predicate or its negation to each of the individuals. For example:

Square(Huey) & not-Red(Huey) & not-Square(Dewey) & not-Red(Dewey) & Square(Louie) & Red(Louie)

In general, if there are n logically independent predicates, then there are 2^n possible combinations of predicates that could be assigned to an individual. If there are m names, each of which could have any of the 2^n combinations, then there are $(2^n)^m$ state descriptions in all.

6 Here's a simple reductio proof: We're given p. Suppose, for the sake of the reductio, that ~(p v q). By DeMorgan's Law, this last sentence implies ~p & ~q, from which ~p follows. But ~p contradicts the premise p. Hence, the assumption ~(p v q) is false, and p v q is true.

7 Thanks to Lance Rips for his valuable comments.

References

Adams, E. (1970) "Subjunctive and Indicative Conditionals." *Foundations of Language* 6: 89–94.

Adler, J. (2002) "Akratic Believing?" *Philosophical Studies* 110: 1–27.

———. (2006) "Confidence in Argument." *Canadian Journal of Philosophy* 36: 225–258.

Ainslie, G. (1992) *Picoeconomics: The Strategic Interaction of Successive Motivational States within the Person* (Cambridge: Cambridge University Press).

Anscombe, G. E. M. (1957) *Intentions* (Oxford: Blackwell).

Axelrod, R. (1984) *The Evolution of Cooperation* (New York: Basic Books).

Belnap, N. (1962) "Tonk, Plonk and Plink." *Analysis* 22: 130–133.

Bennett, J. (2003) *A Philosophical Guide to Conditionals* (Oxford: Oxford University Press).

Boghossian, P. (2000) "Knowledge of Logic" in P. Boghossian and C. Peacocke, eds. *New Essays on the A Priori* (Oxford: Oxford University Press): 229–254.

Bovens, L. (1995) "'P and I Will Believe Not-P': Diachronic Constraints on Rational Belief." *Mind* 104: 737–760.

Bovens, L. and Hartmann, S. (2003) *Bayesian Epistemology* (Oxford: Oxford University Press).

Brandom, R. (1994) *Making It Explicit: Reasoning, Representing, and Discursive Commitment* (Cambridge: Harvard University Press).

Bratman, M. E. (1987) *Intentions, Plans, and Practical Reason* (Cambridge, MA: Harvard University Press).

Burge, T. (1979) "Semantical Paradoxes" *Journal of Philosophy* 76: 169–198.

Campbell, R. and Sowden, L., eds. (1985) *Paradoxes of Rationality and Cooperation*. (Vancouver: University of British Columbia Press).

Carnap, R. (1947) "On the Application of Inductive Logic." *Philosophy and Phenomenological Research* 8: 133–147.

———. (1962) *Logical Foundations of Probability* Second Edition (Chicago: University of Chicago Press).

Carrol, L. (1895) "What the Tortoise Said to Achilles," *Mind* 4: 278–280.

Cherniak, C. (1986) *Minimal Rationality* (Cambridge, MA: MIT Press).

Cohen, L. J. (1977) *The Probable and the Provable* (Oxford: Oxford University Press).

Christensen, D. (2004) *Putting Logic in its Place: Formal Constraints on Rational Belief* (Oxford: Oxford University Press).

Davidson, D. (1980) "How Is Weakness of Will Possible?" in his *Essays on Actions and Events* (Oxford: Oxford University Press): 21–42.

———. (1984) *Inquiries into Truth and Interpretation* (Oxford: Oxford University Press).

Dretske, F. (1970) "Epistemic Operators." *The Journal of Philosophy* 69: 1015–1016.

Duhem, P. (1954) *The Aim and Structure of Physical Theory* (Princeton, NJ: Princeton University Press).

Dummett, M. (1978) "The Justification of Deduction" in his *Truth and Other Enigmas* (Cambridge, MA: Harvard University Press): 290–318.

Elgin, C. Z. (1996) *Considered Judgment* (Princeton, NJ: Princeton University Press).

Elster, J. (1984) *Ulysses and the Sirens*, Second Edition (Cambridge: Cambridge University Press).

Field, H. (1972) "Tarski's Theory of Truth" *Journal of Philosophy* 69: 347–375.

Fine, K. 1975. "Vagueness, Truth and Logic." *Synthese* 30: 265–300.

Fitelson, B. (1999) "The Plurality of Bayesan Measures of Confirmation and the Problem of Measure Sensitivity." *Philosophy of Science* 66: 362–378.

———. (2003) "A Probabilistic Theory of Coherence." *Analysis* 63: 194–199.

Fodor, J. (1983) *Modularity of Mind* (Cambridge, MA: MIT Press).

Frege, G. (1970) "On Sense and Reference" in *Translations from the Philosophical Writings of Gottlob Frege* (Oxford: Blackwell): 56–78.

Garfinkel, A. (1981) *Forms of Explanation* (New Haven, CI: Yale University Press).

Gauthier, D. (1986) *Morals by Agreement* (Oxford: Oxford University Press).

Gibbard, A. (1981) "Two Recent Theories of Conditions" in Harper, W. L., Stalnaker, R. and Pearce G., eds. *Ifs* (Dordrecht: Reidel): 211–247.

Gibbard, A. and Harper, W. (1978) "Counterfactuals and Two Kinds of Expected Utility" in *Foundations and Applications of Decision Theory*, C. A. Hooker et al., eds. (Dordrecht: Reidel).

Ginet, Carl (1980) "Knowing Less by Knowing More." *Midwest Studies in Philosophy V 1980* (Minneapolis: University of Minnesota Press): 151–161.

Glymour, C. (1980) *Theory and Evidence* (Princeton, NJ: Princeton University Press).

Goodman, N. (1965) *Fact, Fiction, and Forecast* (Indiana: Bobbs-Merrill).

Goodman, N. (1983) *Fact, Fiction, and Forecast* (Cambridge, MA: Harvard University Press).

Grice, H. P. (1989) *Studies in the Way of Words* (Cambridge: Harvard University Press).

———. (2001) *Aspects of Reason* R. Warner, ed. (Oxford: Oxford University Press).

Hacking, I. (2001) *An Introduction to Probability and Inductive Logic* (Cambridge: Cambridge University Press).

Hamblin, C. L. (1970) *Fallacies* (London: Methuen).

Hansson, S. O. (2006) "Logic of Belief Revision" in *The Stanford Encyclopedia of Philosophy* Edward N. Zalta (ed.), available at: http://plato.stanford.edu/archives/sum2006/entries/logic-belief-revision/.

Harman, G. (1973) *Thought* (Princeton, NJ: Princeton University Press).

———. (1986) *Change in View: Principles of Reasoning* (Cambridge, MA: MIT Press).

Hempel, C. G. (1965) "Studies in the Logic of Confirmation" in his *Aspects of Scientific Explanation* (New York: The Free Press): 3–46.

Hintikka, J. (1962) *Knowledge and Belief* (Ithaca: Cornell University Press).

Horwich, P. (1990) *Truth* (Oxford: Blackwell).

Hume D. (1977) *An Enquiry Concerning Human Understanding*, E. Steinberg, ed. (Indianapolis: Hackett Publishing Co.).

———. (1978) *A Treatise of Human Nature*, Second Edition, L. A. Selby-Bigge and P. H. Nidditch, eds. (Oxford: Oxford University Press).

Jackson, F. (1975) "'Grue.'" *Journal of Philosophy* 72: 113–131.

———. (1987) *Conditionals* (Oxford: Blackwell).

Jeffrey, R. C. (1983) *The Logic of Decision*, Second Edition (Chicago: University of Chicago Press).

Kaplan, D. (1989) "Demonstratives" in J. Almog, J. Perry, and H. Wettstein, eds. *Themes from Kaplan* (Oxford: Oxford University Press): 481–563.

Kaplan, M. (1996) *Decision Theory as Philosophy* (Cambridge: Cambridge University Press).

Kavka, G. (1983) "The Toxin Puzzle." *Analysis* 43: 33–36.

Keynes, J. M. (1952) *A Treatise on Probability* (London: Macmillan).

Kripke, S. (1975) "Outline of a Theory of Truth." *Journal of Philosophy* 72: 690–716.

———. (1977) "Speaker's Reference and Semantic Reference." *Midwest Studies in Philosophy* 2: 255–76.

———. (1980) *Naming and Necessity* (Cambridge: Harvard University Press).

Kyburg, H. E. (1961) *Probability and the Logic of Rational Belief* (Middletown, CT: Wesleyan University Press).

Kyburg, H. E., and Teng, C. M. (2001) *Uncertain Inference* (Cambridge: Cambridge University Press).

Levi, Isaac. (1980) *The Enterprise of Knowledge* (Cambridge, MA: MIT Press).

Lewis, D. (1971) "Immodest Inductive Methods." *Philosophy of Science* 38: 54–63.

———. (1982). "Logic for Equivocators" *Nous* XIV: 431–441.

———. (1983) "Scorekeeping in a Language-Game" in his *Philosophical Papers: Vol. I* (Oxford: Oxford University Press): 233–249.

———. (1986a) "Causal Decision Theory" (and "Postscript") in his *Philosophical Papers Vol. II* (Oxford: Oxford University Press): 305–339.

———. (1986b) "Probabilities of Conditionals and Conditional Probabilities" (with postscript) in his *Philosophical Papers Vol. II* (Oxford: Oxford University Press): 133–156.

———. (1999) "Why Conditionalize?" in his *Papers in Metaphysics and Epistemology* (Cambridge: Cambridge University Press): 403–407.

Lipton, P. (2004) *Inference to the Best Explanation*, Second Edition (London: Routledge).

Luper, S. (2006) "The Epistemic Closure Principle" in *The Stanford Encyclopedia of Philosophy* Edward N. Zalta (ed.), available at: http://plato.stanford.edu/archives/spr2006/entries/closure-epistemic/.

Makinson, D. C. (1964) "The Paradox of the Preface." *Analysis* 25 205–207.

McGee, V. (1985) "A Counterexample to Modus Ponens." *Journal of Philosophy* 82: 462–471.

Mill, J. S. (1963) "A System of Logic" in *Collected Works of John Stuart Mill Vol. 7–8* J. M. Robson, ed. (Toronto: University of Toronto Press).

Millgram, E. (2001) "Practical Reasoning: The Current State of Play" in his *Varieties of Practical Reasoning* (Cambridge, MA: The MIT Press): 1–26.

Nozick, R. (1969) "Newcomb's Problem and Two Principles of Choice" N. Rescher, ed. in *Essays in Honor of Carl G. Hempel*, (Dordrecht: Reidel)

———. (1981) *Philosophical Explanations* (Cambridge, MA: Harvard University Press).

———. (1993) *The Nature of Rationality* (Princeton, NJ: Princeton University Press).

Parfit, D. (2001) "Rationality and Reasons." *Proceedings of the Aristotelian Society Supplementary* LXXV 1: 195–216.

Parikh, R. (1994) "Vagueness and Utility: the Semantics of Common Nouns." *Linguistics and Philosophy* 17: 521–535.

Peirce, C. S. (1931) *Collected Papers* C. Hartshorne and P. Weiss, eds. (Cambridge, MA: Harvard University Press). Vol. 5: 180–189.

Pollock, J. L. and Cruz, J. (1999) *Contemporary Theories of Knowledge*, Second Edition (Lanham, MD: Rowman and Litlefield).

Popper, K. R. (1959) *The Logic of Scientific Discovery* (New York: Harper).

Priest, G. (1998) "What Is So Bad about Contradictions?" *Journal of Philosophy* 95: 410–426.

Prior, A. (1960) "The Runabout Inference-Ticket." *Analysis* 21: 38–39.

Putnam, H. (1975) "The Meaning of 'Meaning'" in his *Mind, Language and Reality Philosophical Papers Volume 2* (Cambridge: Cambridge University Press): 215–271.

Quine, W. V. O. (1980) "Two Dogmas of Empiricism" reprinted in *From a Logical Point of View*, Second Edition (Cambridge, MA: Harvard University Press): 20–46.

_____. (1969) "Natural Kinds" in his *Ontological Relativity and Other Essays* (New York: Columbia University Press): 114–138.

_____. (1970) *Philosophy of Logic* (Englewood Cliffs, NJ: Prentice Hall).

Rawls, J. (1971) *A Theory of Justice* (Cambridge: Harvard University Press).

Raz, J. (1990) *Practical Reason and Norms*, Second Edition (Princeton, NJ: Princeton University Press).

Reichenbach, H. (1961) *Experience and Prediction* (Chicago: University of Chicago Press).

Richardson, H. (1994) *Practical Reasoning About Final Ends* (Cambridge: Cambridge University Press).

Sainsbury, R. M. (1988) *Paradoxes* (Cambridge: Cambridge University Press).

Searle, J. *Intentionality* (Cambridge: Cambridge University Press, 1983).

Sellars, W. "Empiricism and the Philosophy of Mind" in his *Science, Perception and Reality* (London: Routledge & Kegan Paul, 1963): 127–196.

Smith, E. E., Shafir, E. and Osherson, D. (1993) "Similarity, Plausibility, and Judgments of Probability." *Cognition* 49: 67–96.

Sober, E. (2005) "Bayesianism – Its Scope and Limits" in R. Swinburne, ed. *Bayes's Theorem* (Oxford: Oxford University Press).

Sorensen, R. (2001) *Vagueness and Contradiction* (Oxford: Oxford University Press).

Sperber, D. and Wilson, D. (1986) *Relevance: Communication and Cognition* (Cambridge: Harvard University Press).

Stalnaker, R. C. (1987) *Inquiry* (Cambridge, MA: MIT Press).

Strawson, P. (1952) *Introduction to Logical Theory* (London: Methuen).

Tarksi, A. (1983) "The Concept of Truth in Formalized Languages" in his *Logic, Semantics, and Meta-Mathematics*, Second Edition, J. H. Woodger, trans.; J. Corcoran, ed. and intro. (Cambridge, UK: Hackett) 152–278.

Teller, P. (1973) "Conditionalization and Observation." *Synthese* 26: 218–258.

Tentori, K., Crupi, V., Bonini, N., & Osherson, D. (2007) "Comparison of confirmation measures." *Cognition* 107–119.

Thomson, J. J. (1965) "Reasons and Reasoning" in Max Black, ed. *Philosophy in America* (Ithaca: Cornell University Press): 282–303.

Toulmin, S. (1958) *The Uses of Argument* (Cambridge: Cambridge University Press).

van Fraassen, B. C. (1980) *The Scientific Image* (Oxford: Oxford University Press).

_____. (1984) "Belief and the Will." *Journal of Philosophy* 81: 235–256.

Velleman, J. D. (2000) *The Possibility of Practical Reason* (Oxford: Oxford University Press).

White, R. (2005) "Explanation as a Guide to Induction." *Philosophers' Imprint*, available at: www.philosophersimprint.org/005002/, 5: 1–29.

Williams, B. A. O. (1973) "Internal and External Reasons" in his *Moral Luck* (Cambridge: Cambridge University Press).

Williamson, T. (1994) *Vagueness* (London: Routledge).

Wright, C. (2000) "Cogency and Question-Begging: Some Reflections on McKinsey's Paradox and Putnam's Proof." *Philosophical Issues 10 Skepticism*: 140–163.

Yablo, S. (1993) "Is Conceiveability a Guide to Possibility?" *Philosophy and Pheomenological Research* 53: 1–42.