THE JOURNAL OF PHILOSOPHY

VOLUME LXXXI, NO. 12, DECEMBER 1984

COMPUTATIONAL COMPLEXITY AND THE UNIVERSAL ACCEPTANCE OF LOGIC*

WHEN W. V. Quine says, "better translation imposes our logic" upon the beliefs of any agent we try to interpret, and, furthermore, "the logical truths, or the simple ones, will go without saying; everyone will unhesitatingly assent to them if asked,"[1] it is natural to wonder about the universal acceptance of logic. Let us consider the following version of the thesis: Any rational agent must accept a logic, that is, at least a sound and complete first-order deductive system. I shall argue that the thesis is false under some natural and philosophically important interpretations. The discussion identifies relationships between computational-complexity theory, recent psychological studies of the formal incorrectness of everyday reasoning, and more realistic theories of rationality.

Prima facie, the pattern of complexity-theoretic results in recent years constitutes a kind of practical analogue of the classical absolute unsolvability theorems of the 1930s. The project that emerges is to find the philosophical implications of these results, just as we have been trying to interpret the classical unsolvability results. In particular, if complexity theory in some sense "cuts the computational universe at its joints"—providing a principled basis for a

* This paper was prepared with support from the American Philosophical Society and the National Endowment for the Humanities. Lenore Blum, Charles Chihara, William Craig, Daniel Dennett, Richard Karp, and Barry Stroud generously helped at various stages of the paper. Some of this material appeared in my Ph.D. dissertation, "Belief and Logical Abilities" (Berkeley: University of California, 1977); other material was presented at the University of California, Berkeley, Logic and Methodology Colloquium, February 1982. The paper was read to the Berkeley Philosophy Colloquium, June 1982.

[1] See, respectively, *Word and Object* (Cambridge, Mass: MIT Press, 1960), p. 58; and *Philosophy of Logic* (Englewood Cliffs, N.J.: Prentice-Hall, 1970), p. 102. (Some of Quine's discussion of logic and translation can be construed as entailing only the universal *nonrejectability* of logic. However, the argument below implies that correct translation might attribute rejection of any particular logical truths.)

0022-362X/84/8112/0739/02.00

hierarchy of qualitative distinctions between practically feasible and unfeasible tasks—then we need to examine the idea that, in at least some interesting cases, rationality models ought not to entail procedures that are computationally intractable. Complexity theory raises the possibility that formally correct deductive procedures may sometimes be so slow as to yield computational paralysis; hence the "quick but dirty" heuristics uncovered by the psychological research may be not irrational sloppiness, but instead the ultimate speed-reliability tradeoff to evade intractability. With a theory on nonidealized rationality, complexity theory thereby "justifies the ways of Man" to this extent.

To begin, what is, or would be, a universally accepted logic? At a minimum, the thesis would be that all rational agents accept *some* sound and complete set of laws or axioms and inference rules for first-order logic, as opposed to a claim of the universal acceptance of a particular set of "fundamental" logical laws and rules, or even more strongly, a claim of universal acceptance of all logical truths. The weakest thesis, therefore, could be false either in that: (a) an agent might not accept a complete deductive system; he might have a "cognitive blind spot"; or (b) the agent might accept only an unsound or even inconsistent system; e.g., he might use some rule that did not guarantee preservation of truth in inference, perhaps a "quick but dirty" heuristic; or, more strongly, (c) every law or rule the agent used might be unsound or also inconsistent. I shall deal almost entirely with classical logic. This is not to prejudge the issue of the adequacy of nonstandard logics; the case of classical logic is basic, and the argument should be generalizable to other logics. Whatever one's choice of logic, the prior, and usually unacknowledged, question is whether a sound and complete logic by any standard is in fact the best choice.

I. IDEAL RATIONALITY

Let us examine the concept of a rational agent which is involved in the thesis of the universal acceptance of logic. Some rationality constraint on an agent's cognitive system of beliefs and desires is the most fundamental law of psychology, more basic than any low-level empirical generalization. It is generally recognized in this philosophy of psychology that, for instance, although consistency may be the hobgoblin of small minds, consistency is a condition for having any mind at all: no rationality, no agent. The conventional strategy in the cognitive sciences has been first to adopt a rather extreme idealization of the rationality required of an agent and, then, perhaps, if they are noticed, to explain away departures of real human behavior from the ideal model.

The models of the agent prevalent in decision, game, and economic theory and in philosophy require that the agent be a maximizer of expected utility, that is, that the agent *A* satisfy an ideal general rationality condition, a version of which is: If *A* has a particular belief-desire set, *A* would undertake all and only actions that are apparently appropriate.[2] Here an action is "apparently appropriate" if, according to *A*'s beliefs, it would tend to satisfy *A*'s desires. An agent who is able to choose his actions so well has to have a great deal of logical insight. In particular, he must satisfy an *ideal inference condition*:

> *A* would make all deductively sound inferences from his belief set which are apparently appropriate and would not mistakenly make any unsound ones.

Otherwise, *A* might miss some apparently appropriate actions, for instance.

Now, must an ideally rational agent in this sense accept a logic? For the agent to be able to perform all sound inferences that might turn out to be apparently appropriate and not to make unsound ones, he must meet Cartesian standards of perfection: he must in effect be both infallible and able to have an opinion on anything with respect to logic. Such an agent cannot accept an unsound or incomplete deductive system. If he accepts an unsound system for making some of his inferences, he will not be guaranteed to make only sound inferences, appropriate or otherwise; and if he accepts an incomplete system, he will not be able to make some sound inferences that might turn out to be appropriate. In either case, he will not satisfy the ideal inference condition. Therefore, an ideal agent must accept a sound and complete logic if he is to perform required reasoning by means of a formal deductive system.

II. UNDECIDABILITY

If a rationality requirement is the most basic psychological law, the next most basic fact of our psychology is that we are finite objects. The ideal rationality conditions abstract from this fundamental fact of human existence: we are in the finitary predicament of having fixed limits on our cognitive resources, in particular, on our memory capacity and computing time. The standard model in effect assumes, for such purposes as simplification of theory, that a human being has God's brain; for such an ideal agent much of the deductive sciences would be trivial.

[2] See my "Minimal Rationality," *Mind*, XC, 358 (April 1981): 161–183. (If the requirement that *A* attempt *only* apparently appropriate actions is dropped, a random guesser will satisfy the ideal conditions, given enough time; see the minimal conditions below.)

If we suppose the agent is finite, there is still another cost for the idea of an ideal agent using only formal deductive procedures. For ideal rationality requires more than just use of a sound and complete logic. The agent has to be able to use that logic very well, so well that any given first-order sentence can be formally proved or disproved in a finite number of steps; otherwise, the ideal agent would not be *guaranteed* always to succeed in making all needed sound inferences (e.g., any arbitrary inference he thought his survival depended upon), and also never to make unsound ones. No heuristic procedure for using the deductive system, however good, would suffice unless it was perfect, with no possibility of failure, and hence in fact algorithmic. (Nor could any recursive enumeration procedure suffice by itself, since the agent would wait forever without finding out that some inferences were unsound.) This finitely represented perfect formal ability, of course, would constitute a decision procedure for first-order logic, which Church's theorem demonstrates to be impossible.

Hence, the ideal rationality conditions are very ideal indeed, in that they entail either the most basic practical impossibility—the use of infinite resources—or else a logical contradiction, like a square circle. Of course, the ideal rationality model remains an indispensable simplification of computational reality for many situations—for example, as one norm or "regulative ideal" for evaluating quick and dirty procedures. But some care is required; using the idealization could be a bit as if Hilbert had retained the presupposition of formalism that all number-theoretic truths are formally provable "as a convenient approximation" in the face of Gödel's incompleteness theorem.

But perhaps the agent might accomplish deductive tasks by some entirely nonformal means, for instance, by immediate synthetic a priori intuition of the deductive relations among the propositions he believed. The agent might do this by direct, quasi-perceptual, Gödelian insight into an independent realm of Platonic entities, or by means of his transcendental ego, situated outside of space, time, causality, and so on, as Kant and intuitionists such as Brouwer have identified it. Conformance to the ideal-inference condition through such faculties of intuition may seem little better than doing so by means of an oracle or miraculously perfect luck in guessing; for example, physicalists and those committed to information-processing models of cognition will not be satisfied with even the form of explanations like these. These procedures remain in need of at least the outlines of a scheme of explanation: The alternative procedure has to be such as to guarantee inferential success nonalgorithmically. With intuitive access to a Platonic realm, it is no longer clear that an agent *is* restricted to finite cognitive resources, e.g., of time and space.

III. COMPUTATIONAL COMPLEXITY

In fact, the deductive ability of the ideal agent is even further removed from computational reality. Even where there is no absolute undecidability, a kind of practical undecidability seems to extend further down, to the most basic parts of logic, to the very core of the notion of computation. In some respects, it is as if Church's theorem applied even to the propositional calculus. Of course, a decision procedure exists for tautological soundness—for example, by use of truth tables. But although a tautology decision procedure is in principle possible, it now appears to be inherently "computationally intractable" and, in some sense, to be extremely unfeasible as a practical matter, e.g., to require computations for relatively simple cases that would exceed the capacities of an ideal computer having the resources of the entire known universe. What is the philosophical significance of such intractability?

The above tautology result is in the field of computational complexity, an area that has grown rapidly during the last decade and is yielding practical unsolvability results which may be as interesting in some ways for philosophy as the classical absolute unsolvability results of the 1930s.[3] (Perhaps philosophy has overlooked the field so far because of a tendency to conclude that if a problem is decidable in principle, then it must be trivial, at least for philosophy conceived of as a "pure" nonempirical discipline.[4]) In complexity theory, feasibility of an algorithm is evaluated in terms of whether its execution time grows as a polynomial function of the size of input instances of the problem. If it does (as does any familiar procedure for arithmetical addition, for example), the algo-

[3] Two key papers are S. Cook, "The Complexity of Theorem-proving Procedures," in *Proceedings of the 3rd Annual ACM Symposium on Theory of Computing* (1971): 151–158; and R. Karp, "Reducibility among Combinatorial Problems," in R. Miller and J. Thatcher, eds., *Complexity of Computer Computations* (New York: Plenum Press, 1972): 85–103. A recent review of part of the field is M. Garey and D. Johnson, *Computers and Intractability* (San Francisco: W.H. Freeman Press, 1979). Two easily accessible articles are H. Lewis and C. Papadimitriou, "The Efficiency of Algorithms," *Scientific American*, CCXXXVIII (1978): 96–109; and L. Stockmeyer and A. Chandra, "Intrinsically Difficult Problems," *Scientific American*, CCXL (1979): 140–159.

[4] See S. Kleene's discussion of the concept of a decision procedure in sec. 40 of *Mathematical Logic* (New York: Wiley, 1967) for examples of the relegation of such issues to the applied sciences.

rithm is generally treated as computationally feasible. If it does not and instead increases faster, usually as a exponential function (as does exhaustive search of the game tree in chess, for example), the algorithm is generally regarded as intractable.

Such intractability turns out to a large extent to be independent of how the problem is represented and of the computer model (e.g., random-access or deterministic Turing machine) involved. Just as Turing computability is a formal explication of our intuitive notion of computability, polynomial-time computability can be viewed as one formal specification of a pre-theoretic notion of practical computability. As a first approximation, we can say that complexity theory thereby identifies some of the "natural kinds" of computational difficulty. We shall turn to the question of the "real-world relevance" of complexity theory later; at least important exceptions must be acknowledged to any rule of thumb that equates real-world feasibility with polynomial-time computability.

The ideal agent's procedure for determining whether or not a sentence is a tautological consequence of a set of premises yields a test of whether or not a sentence is truth-functionally consistent. In complexity theory, the latter question is known as the "satisfiability problem." Briefly, the relevant finding here is that the satisfiability problem is a member of the very large and important class of "nondeterministic polynomial time" (NP) problems, which are known to be solvable in polynomial time on a nondeterministic Turing machine, which is allowed to make "guesses" and in effect has an unbounded capacity for some parallel computations. A problem solvable by a nondeterministic Turing machine in polynomial time is solvable by a deterministic machine in exponential time. NP includes P, the class of problems solvable on a standard deterministic Turing machine in just polynomial time.

Most importantly, the satisfiability problem is "NP-complete": *any* NP problem can be efficiently reduced to the satisfiability problem. Each one of the wide variety of known NP-complete problems, numbering in the hundreds, is similarly convertible into any other. In this way, the satisfiability problem is a "universal" NP problem. NP-complete problems have not been proved inherently to require deterministic exponential time; this is the major unanswered question, a "Goldbach's conjecture", of the field, equivalent to the question whether NP $\neq$ P in that some NP problems are not in P. However, NP-complete problems are generally regarded as computationally intractable in this way, since only exponential-time deterministic algorithms for any of them are known, and because if they were not, so many important problems that have long resisted practical solution (such as the "traveling-salesman problem"[5]) would then all turn out to be tractable.

Thus, the ideal agent's perfect capacity even just to make all tautological inferences is the case *par excellence* of a problem-solving capacity that is strongly conjectured to require computationally intractable algorithms. Of course, a "quick but dirty" heuristic procedure for tautological inference will not necessarily yield such apparent exponential explosion of computation—presumably, that is how actual fallible human beings manage, as we shall see. But again, nonalgorithmic procedures would not suffice for the Cartesian perfection of the ideal agent, since *ex hypothesi* they cannot be guaranteed to work in all cases. A surprisingly small and basic fragment of the ideal agent's deducing ability seems by itself to require, for just a finite set of simple cases, resources greater than those available to an ideal computer constructed from the entire universe. There is another layer of impossibility between the idealization and reality, not merely minor exceptions.

IV. MINIMAL RATIONALITY

We can therefore say that, although use of an ideal rationality model is an understandable motivation for arriving at the thesis of universal acceptance of logic, some other argument still is needed for that thesis. Although ideal and more realistic models ought to coexist, for some purposes the idealization strategy seems an overreaction to the "no rationality, no agent" point. The alternative approach is to begin with a somewhat less idealized, more realistic model, of *minimal* rationality, where the agent's ability to choose actions falls between randomness and perfection. A minimal general rationality condition would be: If *A* has a particular belief-desire set, *A* would undertake some, but not necessarily all, actions that are apparently appropriate. For an agent to satisfy this condition, he must have some, but not ideal, logical ability—that is, he must satisfy a *minimal inference condition*:

> *A* would make some, but not necessarily all, sound inferences from his belief set which are apparently appropriate.

(*A* must also *not* attempt enough of the actions that are apparently inappropriate, and inferences that are unsound or apparently inappropriate.)

A useful feature of this less idealized model of rationality is that,

[5] The traveling-salesman problem, an NP-complete network design problem of considerable practical interest in operations research, is: given a set of cities on a map and all intercity distances, construct the optimum tour, i.e., the shortest round-trip route connecting all the cities by intercity links.

as we shall see, it provides a philosophical framework for relating two areas of significant research during the last decade. One is the field of computational complexity just sketched. The other encompasses the many recent psychological experiments that suggest surprisingly ubiquitous use of prima facie sub-optimal "heuristic strategies", rather than formally correct procedures, in everyday intuitive reasoning.[6] Although each of these areas has arisen independently of the others, there seems to be a fundamental connection: (a) Complexity theory provides a principled basis for considering that human beings (indeed, any computational entities) may not be able to perform some very simple reasoning tasks in ways that are guaranteed to be correct. (b) The psychology of "irrationality" suggests how we can do these tasks, by showing something of how we in fact do them—by means of the "quick but dirty" heuristics. (c) The ideal rationality models are at best silent on the normative status of use of these heuristics; the minimal rationality model, to begin with, permits us to acknowledge the basic platitude that human beings are in the finitary predicament, and so *ought* to use some such heuristics—according to this conception, formally incorrect heuristics need not in fact be irrational at all. They are not just inadvisable or unintelligible sloppiness, because they are a means of avoiding computational paralysis while still doing better than guessing.

The increasing interest in computational complexity and also in psychological heuristics makes it important to establish the status of claims of human (or even inherent computer) alogicality or illogicality. In particular, are the claims somehow a priori incoherent and so not a matter open for empirical study, as the rationality idealizations—and even the usual "charity principles"—imply? We thereby return to the issue of the universal acceptance of logic. Given the limitations of the concept of the ideal agent, the main question has now become: If a supposed cognitive system qualifies as minimally rational, is there any sense in which it must include a logic?

To determine in what sense, if any, satisfaction of the minimal rationality conditions implies acceptance of a deductive system, we must ask: What is accepting (or believing) a logical law or rule? Briefly, let us distinguish between strong and weak acceptance of

[6] For a review of some of their own basic work by two major contributors to the field, see A. Tversky and D. Kahneman, "Judgment under Uncertainty: Heuristics and Biases," *Science*, CLXXXV (1974): 1124–1131. A recent overview is R. Nisbett and L. Ross, *Human Inference: Strategies and Shortcomings of Social Judgment* (Englewood Cliffs N.J.: Prentice-Hall, 1980).

logic. Assent to a logical law, mere lip service, is not enough to constitute strong belief in the law. Assent is neither sufficient nor necessary, although it is one type of evidence for such acceptance. In addition, acting appropriately for, or reasoning in accordance with, a logical law is not enough to constitute such belief. For instance, a sound argument is "in accordance with" *every* valid sentence, in the sense that the argument's conclusion also follows from the premises conjoined with any of these validities; there is then no distinction with regard to accepting logic between idiot and super-savant.

As Donald Davidson has emphasized, a belief must be part of an agent's reason for a decision.[7] Causal efficacy, the "right" role in the decision-making process, also is required here; the minimal agent must actually use the law as a premise in some (not necessarily all) of the practical reasoning, conscious or unconscious, by which he would select apparently appropriate actions (There can be important "generate and test" interplay between heuristic reasoning in the context of discovery, and logic as *post hoc* tribunal in the context of proof.) The key notion in turn, therefore, is that of "using a logical law or rule" (the related notion of "following a rule" has received much attention[8]).

In contrast, to accept or believe a logical law weakly is merely to be usefully (or instrumentalistically) described as using the law; it may be clear that the agent is not in fact using the law at all. This appears to be the sense in which Daniel Dennett says of adaptively behaving creatures from another planet, "in virtue of their rationality they can be supposed to share our belief in logical truths," and further, of "mice and other animals, in virtue of their being intentional systems," "whether or not the animal is said to *believe* the *truths* of logic, it must be supposed to *follow* the *rules* of logic."[9]

Thus, a person might strongly accept a logic—a small set of simple axioms and inference rules from which all logical truths could "in principle" be derived. But such strong acceptance of a complete deductive system for first-order logic is not the same as strong acceptance of "the theory of first-order logic": actual appropriate use of each of the infinitely many assertions derivable by means of those axioms and rules. However, although the latter is not possible for a realistic or minimal agent, the agent can resolve to accept, or be

[7] See, for example, "Psychology as Philosophy," in *Essays on Actions and Events* (New York: Oxford, 1980).

[8] Cf. Wittgenstein's discussion in *The Blue Book* (New York: Harper & Row, 1958).

[9] "Intentional Systems," in *Brainstorms* (Montgomery, Vt.: Bradford Books, 1978), pp. 9, 11.

committed to accepting, these truths. Also, a person can endorse a deductive system—for instance, as an object language for the relatively restricted technical purposes of metamathematics. The more limited the use of the system—the more it is preached as a norm on Sunday, but not practiced the rest of the week—the more such endorsement tends to fall below strong acceptance.

V. PRACTICAL ADEQUACY OF A LOGIC

The question of universal acceptance of logic now becomes, Must a minimally rational agent accept a logic either strongly or weakly? The interesting issue is whether an agent's satisfaction of the minimal rationality conditions implies his strong acceptance of a sound and complete deductive system. (The argument below can also be adapted for the weak sense of 'accept'.) The question needs further sharpening: One might argue[10] that, though an agent must be able to make some sound inferences—that is, must have some deductive ability—he does not have to be able to make any *particular* inferences, even those which normal human beings find the most obvious. But even if this is true, it does not exclude universal acceptance of logic. It might still turn out that any agent must strongly accept some complete set of valid laws and sound rules; it would just be that agents do not have to accept the laws and rules normal humans do—for instance, those normal humans find obvious. Our question is therefore, Must an agent strongly accept *any* set of valid laws or sound rules that constitute a complete deductive system, much less particular obvious laws or rules? It seems that an agent can have the deductive ability required by the rationality conditions without strongly accepting, i.e., actually using sometimes, even one such law or rule. (We restrict consideration to verbally formulated beliefs.)

We need one more distinction. I shall say that a deductive system is *metatheoretically* adequate if it is sound (and therefore consistent) and complete. In the first paragraph of "The Justification of Deduction", Michael Dummett asserts, "Failure of soundness yields a situation which must be remedied. Failure of completeness cannot always be remedied; a remedy is, however, mandatory whenever it is possible."[11] Dummett accurately describes adherence to

[10] As I have in my "Feasible Inferences," *Philosophy of Science*, XLVIII, 2 (June 1981): 248-268. See also sec. 5 of that paper for an argument against the claim that being able to make particular "obvious" inferences, by whatever means, is *constitutive* of understanding the logical constants involved in those inferences; the argument would apply in particular against the assertion that strong acceptance—that is, actual use—of the corresponding obvious valid laws or sound rules is required for an agent to qualify as understanding the logical constants involved.

[11] In *Truth and Other Enigmas* (Cambridge, Mass.: Harvard University Press, 1978), p. 290.

such an absolute requirement as "the standard practice of logicians" in constructing and justifying a formal logical theory. However, the metatheoretic adequacy of a deductive system must be explicitly distinguished from its *practical* adequacy: here, its adequacy for accomplishing the deductive tasks required of a minimally rational agent. If one assumes that any possible agent must be ideally rational, it is easy to overlook the difference between the two types of adequacy. But with a minimal rationality model, the overlap of the two types of adequacy becomes much less salient.

What is the relation between metatheoretic and practical adequacy? I shall point out that practical adequacy does not require metatheoretic adequacy, that the former is sometimes preferable to the latter, and that the former may sometimes not even be compatible with the latter. The very quickest possible, but least reliable, way of performing a deductive task is just to guess the answer. We know that a minimal agent does not have to be a perfect logician, but the agent could not accomplish his required sound inferences (while avoiding enough unsound ones) just by a series of lucky guesses. There are, however, other ways to improve above chance the odds of selecting conclusions that follow from premises besides using a sound and complete deductive system. The agent might use what is in effect a better than random, but not perfect, gambling strategy for identifying sound inferences. Though such a rule of thumb would not always succeed, it might work sufficiently often to reach the break-even point of satisfying minimal rationality requirements. I shall argue later that this type of strategy may be indispensable if computational intractability is to be avoided.

VI. AGAINST METATHEORETIC ADEQUACY

The concept of such a strategy suggests, to begin with, that it is at least possible for a logically competent agent to have one or more "logical blind spots" which are the result of his exclusively using an incomplete deductive system (whenever he does use a deductive system). Given the difficulties for observer—and agent—in determining the agent's nonconscious cognitive processes, let us consider the overt and explicit steps written out by a person performing a formal derivation (or alternatively, the core dump of a computer running a theorem-proving program). As an uninteresting example, the agent might use a conventional textbook natural-deduction system of independent rules, with a *modus tollens* rule that has a clause that excludes its application to formal sentences with more than 1,000 logical constants. Similarly, it is possible for an agent to perform all required sound inferences by means of an unsound, or even inconsistent, system. Frege's axiomatization of logical theory in *The Fundamental Laws of Arithmetic* and Quine's in *Mathematical Logic* were both inconsistent in ways that did not reveal themselves to many who had used each axiomatization extensively. These two examples suggest that first-order deductive systems can correspondingly be inconsistent in ways that do not yield too many—indeed, any—unsound inferences for the range of deductive tasks required of an agent. And in fact all the early formulations of the substitution rule for the predicate calculus are reported to have been unsound.[12] (Of course, if a deductive system is inconsistent it is complete, but this is no longer metatheoretic virtue.)

Furthermore, an agent who satisfied the minimal rationality condition could use exclusively a deductive system in which *all* axioms were invalid and *all* inference rules were unsound. A natural-deduction system corresponding to a standard textbook one, but composed entirely of unsound rules can easily, if uninterestingly, be constructed. For example, to the original *modus tollens* rule a clause is added: "When one of the premises contains more than 1,000 logical constants, the set of premise-numbers of the line on which the conclusion occurs should be empty; otherwise the premise-numbers are as usually specified." A similar premise-number clause can be added to each of the other rules. Or the original other rules and the new *modus tollens* rule can just be conjoined as a single rule; such a matter of individuating rules seems arbitrary.

The agent would not use the usual shortcut "theorem" rule which permits entering in a derivation a previously proved theorem with an empty set of premise-numbers. The claim here is just that this agent *can* use exclusively this set of unsound basic rules. The agent might happen to be uninterested in using that set to deduce "vacuous" valid sentences; perhaps, as empirical studies indicate for normal human beings, he has difficulty reasoning so abstractly. As proponents of the naturalness of natural-deduction systems often point out, outside of logic courses people rarely seem to use, or at least to cite, logical validities.

The unsound inferences permitted by this system would be performed relatively rarely because they would arise only under a restricted range of conditions: they involve very complex sentences, or might be otherwise unintuitive or difficult for the agent to perform. We know that an agent cannot perform all inferences—in particular, the more complex ones—anyway; so the unsoundness of this system need not detract *at all* from the agent's rationality. We

[12] S. Kleene, *op. cit.*, fn, p. 107. See A. Church, *Introduction to Mathematical Logic*, vol. 1 (Princeton, N.J.: University Press, 1956), pp. 289/90.

conclude that metatheoretically adequate deductive systems are not the *only* way to achieve practical adequacy. One cannot argue that any possible rational agent must accept logic.

Furthermore, a stronger point against metatheoretic adequacy seems to hold: that in important cases it is antagonistic to practical adequacy. A metatheoretically inadequate system could be superior to any metatheoretically adequate one for the practical purposes of accomplishing an agent's everyday deductive tasks, just as inconsistent naive set theory is often more convenient than one of the consistent axiomatizations. In such a situation, insisting upon use of a metatheoretically adequate system would itself be unreasonable, like trying to use the more correct but hopelessly unwieldly relativity physics instead of classical mechanics for engineering calculations in designing a bridge. Even outside of practical contexts empirical theories are often recognized to be idealizations that are only approximations of reality and apply satisfactorily only over limited ranges of the parameters involved; the kinetic-molecular theory of gases is a standard example of an idealization that is employed because it is much more manageable than more correct theories, e.g., that do not assume molecules are perfectly dimensionless spheres.[13]

Indeed, as mentioned earlier, much evidence has recently emerged indicating that in a remarkably wide range of conditions human beings do not in fact use formally correct procedures in everyday nondeductive reasoning. And occasionally researchers in this field have pointed out, in effect, that use of such quick but dirty heuristics in "applied" as opposed to "pure" situations may be a reasonable speed-accuracy tradeoff.[14] There is also a separate tradition of empirical research suggesting that people do not use formally correct procedures in simple deductive reasoning. In addition, some of my own recent empirical studies indicate that subjects use a "prototypicality heuristic" in deductive reasoning, a set of shortcut strategies that exploit structuring of concepts in terms of prototypes, or best examples, of the concepts; we seem to extend the "context of discovery" in this way into the "context of proof." Furthermore, some of the evidence suggests that using this formally incorrect procedure is in fact rational, in that it pays off with lower error rates. These last findings are significant because they identify a connection between deductive-reasoning heuristics and the important recent research on prototype models of mental rep-

[13] The question of adopting convenient but in some sense unsound inference rules in fact sometimes is raised in devising natural-deduction systems for introductory logic texts; see, for example, Mates, *op. cit.*, p. vii.

[14] See, for example, the last chapter of Nisbett and Ross, *op. cit.*

resentation.[15] Finally, it is worth recalling in this context the widespread occurrence of the simplest classical semantic and set-theoretic antinomies in our conceptual scheme, from the foundations of mathematics to ordinary discourse; this may be another symptom of our use of formally incorrect deductive procedures.

VII. PRACTICAL PARALYSIS

An even stronger point may hold than just that metatheoretically inadequate systems appear to be preferable sometimes to metatheoretically adequate ones. Some of the recent research on computational complexity raises the possibility that metatheoretic adequacy may in important ranges of cases be entirely incompatible with practical adequacy, both for some of the "pure" purposes of the deductive sciences and for the "applied" purposes of maintaining an agent's minimal rationality. Let us begin with an argument of Michael Rabin's[16] for the introduction of a notion of probabilistic proof in mathematics. Of course no decision procedure is possible even in principle for all of elementary number theory, but even in-principle decidability can sometimes be of very limited value. Consider a result of Albert Meyer and Larry Stockmeyer's[17]: Although the set of theorems of a formal system for the weak monadic second-order theory of successor (WS1S), a fragment of elementary number theory, is decidable in principle, its decidability seems extremely unfeasible in practice. The problem requires not just exponential time, but "super-exponential" time. To prove theorems of only 617 symbols or less would require a network with so many boolean elements that, even if each were the size of a proton, the machine would exceed the size of the entire known universe. In effect, the moral Rabin drew from the pattern of such complexity-theoretic results is that, to avoid the problem of unfea-

[15] An early (and controversial) study that suggested subjects use a kind of global impression or "atmosphere" of the logical form of the inference was R. Woodworth and S. Sells, "An Atmosphere Effect in Formal Syllogistic Reasoning," *Journal of Experimental Psychology*, XVIII (1935): 451-460; see also P. Wason and P. Johnson-Laird, *Psychology of Reasoning* (Cambridge, Mass.: Harvard, 1972), ch. 10. For a review of research on prototypicality, see E. Rosch, "Human Categorization," in N. Warren, ed., *Studies in Cross-cultural Psychology*, vol. 1 (New York: Academic Press, 1977). On the use of a prototypicality heuristic in deductive inference, see my "Prototypicality and Deductive Reasoning," *Journal of Verbal Learning and Verbal Behavior* (October 1984).

[16] See "Theoretical Impediments to Artificial Intelligence," in J. Rosenfeld, ed., *Information Processing 74* (Amsterdam: North Holland, 1974). On Rabin's concept of a probabilistic algorithm, see his "Probabilistic Algorithms," in J. Traub, ed., *Algorithms and Complexity* (New York: Academic Press, 1976). Rabin's argument is described in G. Kolata, "Mathematical Proofs: The Genesis of Reasonable Doubt," *Science*, CXCII (1976): 989/90.

[17] See A. Meyer, "Weak Monadic Second-order Theory of Successor Is Not Elementary-Recursive," in A. Dold and B. Eckmann, eds., *Lecture Notes in Mathematics*, no. 453 (New York: Springer-Verlag, 1975); and Stockmeyer and Chandra, *op. cit.*

sibly long proofs, mathematicians sometimes should make the ultimate speed-accuracy tradeoff: relax the metatheoretic requirement of consistency, even where it is in principle satisfiable, to evade practical paralysis. Rabin recommended, and devised, methods of probabilistic proof which do not guarantee truth, but for which the probability of error can be determined to be, e.g., one in a billion.[18]

Rabin's strategy might be compared with undoing Descartes's bargain; Descartes sought apodeictic certainty, but the cost is recognized to have been epistemic paralysis. We can extend the point from methodology of the deductive sciences to our concern about fundamental constraints on human cognition. There are now two possible extremes for dealing with, e.g., the problem of determining tautological consequence. Just guessing, the quickest but dirtiest procedure, is too dirty for even minimal rationality, since odds of success are chance. The other extreme, a decision procedure, is the most reliable but also seems too slow; it is perfectly infallible, but, as noted earlier, it is probably computationally intractable, which might be too slow for even minimal rationality. Therefore, a compromise between the two extremes seems needed to yield sufficient deductive capacity for minimal rationality

Various tradeoff strategies are in fact prevalent in computer science in dealing with problems that are found to be computationally intractable. Standards are lowered and "heuristic algorithms" and "approximation algorithms" are sought for the problems instead of perfect optimization algorithms.[19] There is more than one type of compromise with perfect algorithmhood which might evade the apparent intractability of decision procedures for even just tautologous consequence. One might, for instance, use a metatheoretically adequate deductive system, but avoid worst cases by restricting its application to simpler special cases—sets of premises and a possible conclusion that are small enough so that the exponential explosion of operations is not severe. Therefore, the agent could never attempt to make an inference from, or to test for consistency, his entire belief set, or even a large portion of it. The cost, and the eventual limitation, is that the agent would then exhibit the most rationality "locally", within certain neighborhoods of his beliefs, and would be particularly weak on inferences and consistency involving beliefs distributed between such subsets. (In fact, a fundamental feature of human belief systems is that they are

[18] Hilary Putnam has also argued, in a Quinean vein, that "quasi-empirical" methods—resembling those employed in evaluating the plausibility, for instance, of highly theoretical statements in physics—have always been important in mathematics outside of the domain of formal proof. See "What Is Mathematical Truth?", in *Philosophical Papers,* vol. 1 (New York: Cambridge, 1975).

[19] See, e.g., Garey and Johnson, *op. cit.*, ch. 6.

"compartmentalized" in this way; thus, an important rationale for this structuring is as a strategy for evading intractability.[20]) Another, simpler strategy would be: Given any case, employ the metatheoretically adequate system; if no answer results within some fixed time limit, just abandon the attempt and flip a coin to pick an answer.

VIII. REAL-WORLD RELEVANCE

These strategies, however, require that computationally manageable cases involving sufficiently large belief sets be sufficiently frequent to avoid *de facto* computational paralysis. We must therefore turn to the issue of the "real-world relevance" of complexity theory. It should be emphasized that usual complexity measures are not average-case estimates. The typical theorem stating that a problem is computationally complex is of a worst-case form: Given any algorithm for deciding each instance of the problem, each of an infinite number of cases requires exponential time.

This still leaves in limbo an infinite number of *other* cases of the problem. Which of them, if any, requires exponential time? Perhaps every instance of the problem takes exponential time; even then, the exponential blowup might be so slow that the computational cost is not severe for small cases of the problem. Or instead, perhaps no real-world relevant case—of less than colossal size—might require "serious" exponential time; the exponential cases might be of an input size that nothing of human-scale computational resources could ever even encounter. Or again, the problem's complexity profile might fall messily between these two extremes. Therefore information is needed on the "density", or population distribution, of the hard cases. This includes, for example: (a) Do they arise "early" (for cases of about the same order of size as the shortest decision algorithm for the problem)? (b) Do they arise "often", that is, within the population of *interesting* cases; this requires an understanding of which instances are interesting, which of course will be relative to particular goals. (c) How severe is the exponential explosion, when it does occur?

The probability distribution of relevant worst cases is presently not well understood.[21] Indeed, algorithms for linear programming

[20] See my "Rationality and the Structure of Human Memory," *Synthese*, LVII, 2 (November 1983): 163-186.

[21] For an investigation of probability-distribution, as opposed to worst-case, analyses, see R. Karp, "Probabilistic Analysis of Some Combinatorial Search Algorithms," in J. Traub, ed., *op. cit.* Also relevant for the satisfiability problem is S. Mahaney, "Sparse Complete Sets for NP: Solution of a Conjecture by Berman and Hartmanis," *21st IEEE Symposium on Foundations of Computer Science* (1980): 54-60.

are a well-known counterexample to the assertion: "An algorithm is in fact practically unfeasible if and only if it requires exponential time." On the one hand, the Simplex linear programming algorithm has been proved to require exponential time. Yet for decades the Simplex algorithm has been found very usable in practice—for the population of problems of interest to its users. A recent "empirical" study of running times of the algorithm for actual problems in the banking, steel, and oil industries confirms this; and Steven Smale has just proved that the exponential cases are in a sense rare. On the other hand, the "Khachian" algorithm requires only polynomial time, but its typical running time seems much worse than that of the Simplex algorithm, because its polynomial bounds are so high.[22] The connection of exponential time with in-practice unfeasibility thus needs to be interpreted with some care.

Nonetheless, workers in many areas of computer science certainly continue to accept this connection as a very useful rule of thumb.[23] And some rough estimates do not suggest optimism that methods like the above compartmentalizing and "give up and guess" strategies are by themselves sufficient to avoid intractability in the management of a human belief system. Even decidability of just the monadic predicate calculus (and of some other decidable subclasses of the full predicate calculus) is known to require nondeterministic exponential time.[24]

And, at least as food for thought, it is worth again considering testing for tautological consequence, only a very small part of the general problem here, by means now of the truth-table method (even the more efficient known test procedures such as Wang's algorithm or the resolution method still require as much time in the worst cases). Given the difficulties in individuating beliefs, it is not easy to estimate the number of logically independent atomic propositions in a typical human belief system, but 138 seems much

[22] For the proof that the Simplex algorithm requires exponential time, see V. Klee and G. Minty, "How Good Is the Simplex Algorithm?" in O. Shisha, ed., *Inequalities-III* (New York: Academic Press, 1972). The "empirical" study of running times is in E. McCall, "Performance Results of the Simplex Algorithm for a Set of Real-world Linear Programming Models," *Communications of the Association for Computing Machinery*, xxv (1982): 207-212. Smale's proof has been reported in *Science*, CCXVII (1982): 39. A short review of the so-called "Khachian" algorithmn: L. Lovacs, "A New Linear Programming Algorithm—Better or Worse than the Simplex Method?" *Mathematical Intelligencer*, II (1980).

[23] See, for example, pp. 8/9, Garey and Johnson, *op. cit.*

[24] For a review of these results, see H. Lewis, "Complexity of Solvable Cases of the Decision Problem for the Predicate Calculus," *Proceedings of the 19th IEEE Symposium on Foundations of Computer Science* (1978): 35-47.

too low—too "small-minded". Yet suppose that each line of the truth table for the conjunction of all these beliefs can be checked in the time a light ray takes to traverse the diameter of a proton, an appropriate cycle time for an ideal computer. At this maximum speed, a consistency test of this very modest belief system would require more time than the estimated twenty billion years from the dawn of the universe to the present. Quinean or Davidsonian charity requirements that a translation be readjusted if it yields an inconsistent belief set seem particularly unrealistic in this light.

Furthermore, it is important to note that some exponential-time problems have been proved to have hard cases of small size (indeed, similar proofs have been emerging for classical absolutely unsolvable problems[25]). As noted above, deciding WS1S sentences of just several hundred symbols is known to require more space (and time) than there is in the known universe. Also, Fischer and Rabin have showed that Presburger arithmetic requires exponential time and, furthermore, that the exponential explosion of proof length sets in early, for sentences of the same order of size as the decision algorithm.[26]

Nonetheless, it is quite easy to show that an early onset of computational complexity—that is, for cases small enough to be humanly relevant—is not an inherent feature of intractability: Given any intractable problem with early-onset complexity, one can always construct another problem with complexity onset only for cases larger than a particular given size.[27] The significant implication here of this point is that "counting the horse's teeth"—empirically observing which are the difficult and interesting cases of a

[25] Undecidability for elementary arithmetic is now known not to arise only for extremely complex and mathematically uninteresting sentences: See J. Jones, "Three Universal Representations of Recursively Enumerable Sets," *Journal of Symbolic Logic,* XLIII (1978): 335-351, for an undecidable, unabbreviated sentence (based on the solution of Hilbert's Tenth Problem) of about 100 symbols in length; and his "Universal Diophantine Equation," *Journal of Symbolic Logic,* XLVII, 3 (September 1982): 549-571, for the basis for constructing still shorter sentences. See also G. Chaitin, "Information-theoretic Limitations of Formal Systems," *Journal of the Association for Computing Machinery,* XXI (1974): 403-424. In addition, a "mathematically simple and interesting" theorem (an extension of the finite Ramsey theorem) is not provable in Peano arithmetic; see J. Paris and L. Harrington, "A Mathematical Incompleteness in Peano Arithmetic," in J. Barwise, ed., *Handbook of Mathematical Logic* (New York: North-Holland, 1977). The present paper further motivates study of the "density", or distribution, of undecidable sentences.

[26] M. Fischer and M. Rabin, "Super-exponential Complexity of Presburger Arithmetic," in *Complexity of Computation, SIAM-AMS Proceedings,* VII (1974): 27-41.

[27] For example, as noted, the set of theorems of Presburger arithmetic has early-onset complexity. However, the set of Presburger arithmetic theorems, each of length more than 10,000 symbols, can be quickly decided for any case of length less than 10,000 symbols; the algorithm will just count the number of symbols in any sentence to be decided and immediately reject it if that length is less than 10,000.—Another question concerns the "naturalness" of such truncated problems.

problem—will often be unavoidable; much of the "real-world" complexity structure of many intractable problems is at least presently a hybrid question, to be approached in the manner of McCall's study of the Simplex algorithm cited earlier. In particular, artificial intelligence workers on automatic theorem proving have uniquely valuable data about what are the "interesting", frequently occurring, cases of a given intractable problem, and which of them in fact require large expenditure of resources. And similar information can correspondingly be obtained for human deductive reasoning. That is, in the hypothetico-deductive manner, we can treat the issue of whether there is real-world relevant complexity at least as an empirical working hypothesis, suppose it is true, and see how well its implications are supported by observations.

In fact, the rather wide range of recent evidence mentioned earlier confirms that people actually do not use formally correct methods in their intuitive reasoning. Instead, they seem to use procedures in nondeductive everyday reasoning such as Kahneman and Tversky's "representativeness" and "availability" heuristics, and related "prototypicality heuristics" in deductive reasoning. Thus, the most fundamental "empirical" suggestion of complexity theory, that algorithmic methods of accomplishing even some very simple deductive tasks are likely to be intractable, would provide a unifying framework for explaining why we use such heuristics instead. It constitutes the ultimate "justificaton of the ways of Man" here. For a plausible conjecture to begin with is that our quick and dirty shortcut strategies are required to avoid intractability. To the constraint that information-processing models of cognition should be finitary, we could then add that, in at least some interesting cases, they also should not entail certain classes of computationally complex processes. The possibility needs to be explored further that, to a considerable extent, the only way human reasoning (or any agent's) can evade practical paralysis is by not using metatheoretically adequate deductive systems.[28] Ideal rationality requirements seem to exclude even entertaining such a possibility as more than unimportant exceptions to a rule, but a theory of minimal rationality provides a principled basis for relating complexity theory and the psychological studies in this way.

[28] Correspondingly in artificial intelligence, the algorithmic approach of seeking classical decision procedures, or even just complete proof procedures, for theoremhood needs to be reevaluated in light of the problem of combinatorial explosion of branchings in proofs. For a brief discussion of the issue, see A. Newell and H. Simon, "Computer Science as Empirical Inquiry," *Communications of the Association for Computing Machinery,* XIX (1976): 113-126. See also Rabin, "Theoretical Impediments to Artificial Intelligence," *op. cit.* An extensive review of practical limitations on theorem-proving programs (and program verifiers) still is needed.

The familiar debate about how to choose between competing logics, e.g., which is more "natural", presupposes as an absolute requirement that we ought not even to consider a logic that is metatheoretically inadequate. But we can now see that the commonplace that Dummett, for example, began with, that metatheoretic adequacy is mandatory whenever "in principle" possible, needs at least to be carefully restricted. If one is engaged in a metatheoretical investigation of a deductive system, rather than generally using the system, Dummett's assertion may be correct. An agent can endorse the use of a metatheoretically adequate system, reason *about* it, and even really use it in some limited contexts. But if the system is actually to be used significantly, metatheoretic inadequacy may be reparable only on pain of intractability; it would then be irrational even to try to adopt exclusively a metatheoretically adequate system, since that would preclude successful reasoning. It would be like insisting upon the perfectionism of the Cartesian methodology of universal doubt, with its resulting cognitive paralysis.

Thus, in contrast to Alfred Tarski's remark, "the appearance of an antinomy is for me a symptom of disease,"[29] there is at least some truth to Wittgenstein's earlier image in his account of the (non)significance of the paradoxes, "a contradiction is not a germ which shows a general illness."[30] The contradiction need not *entirely* vitiate the system. Indeed, this paper has gone further, proposing that such inconsistency may be downright healthy. The moral regarding the thesis of the universal acceptance of logic is that once we begin to take into account at least the most fundamental facts of an agent's psychological reality, that thesis, like a number of other rationality idealizations, seems wrong, and interestingly so. Contrary to the usual charity principles, not only is acceptance of a metatheoretically adequate deductive system not transcendentally indispensable for an agent's rationality, but in important cases it is inadvisable and perhaps even incompatible with that rationality.

CHRISTOPHER CHERNIAK

University of Pennsylvania

[29] "Truth and Proof," *Scientific American*, CCXX (1969): 63-77.

[30] *Wittgenstein's Lectures on the Foundations of Mathematics: Cambridge, 1939*, C. Diamond, ed. (Ithaca, N.Y.: Cornell, 1976), p. 211.