AI Magazine Volume 4 Number 3 (1983) (© AAAI)

What is Rational Psychology?
Toward a modern mental philosophy

Jon Doyle

Computer Science Department
Carnegie-Mellon University
Pittsburg, PA 15213

Abstract

Rational psychology is the conceptual investigation of psychology by means of the most fit mathematical concepts Several practical benefits should accrue from its recognition

SOME PROBLEMS closely associated with those of artificial intelligence and cognitive science seem unduly neglected in light of the possible benefits of their investigation. These are the problems of investigating theories and techniques of natural and artificial psychologies by means of the most fit mathematical concepts. The term "rational psychology" labels this investigation. Rational psychology should not be confused with logic-based presentations of artificial intelligence. While investigations based on mathematical logic are relatively familiar and certainly useful, using only that portion of mathematics to characterize psychologies presupposes that psychological questions are fundamentally logical. That presupposition is not necessary for the development of an exact science of mind. To urge the broader view, the following briefly explains the idea of rational psychology, places it among its associated fields, and indicates some of its likely benefits.

This research was supported by the Defense Advanced Research Projects Agency (DOD), ARPA Order No 3597, monitored by the Air Force Avionics Laboratory under Contract F33615-81-K-1539 The views and conclusions contained in this document are those of the author, and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the Government of the United States of America

Rational Psychology

Rational psychology is a part of mathematics, the conceptual investigation of psychology. "Rational" here indicates psychological investigations based on reason alone, rather than on experiment, engineering, or computation, the rational analysis of the concepts and theories whose applicability and feasibility are studied in experimental, engineering, and computational projects Rational psychology is not the study of rational agents, but instead the mathematical approach to the problems of agents and their actions, whether these agents and actions are themselves thought rational or irrational. The name stems from the rational mechanics of Newton, and is merely adaptation to the realm of mental philosophy of the principles, aims, and methods found in his natural philosophy (Truesdell 1958) Although I contrast rational psychology with other disciplines, the term

This paper abbreviates the earlier version of Doyle (1983) I thank Joseph Schatz, Johan de Kleer, Gerald Sussman, Jaime Carbonell, Allen Newell, Merrick Furst, Raymond Reiter, John MacNerney, and Robert Laddaga for several discussions I have borrowed heavily from the writings of Clifford Truesdell Compare also Hartmanis (1981), Minsky (1962), and Nilsson (1980)

is not meant to exclude others but to highlight a common project occurring in specialized and isolated manifestations. It is not meant merely to agglomerate numerous disciplines, nor to prevent specialization. The aim is instead to reset the common foundations of mental fields to make the unity apparent mathematically while aiding the prosecution and communication of specialized inquiries.

This enterprise involves a different conception of what is meant by "mind," "mental," and "psychology" than that common in the existing mental sciences. In the following, a psychology is merely a specification of the structure and behavior of some agent, and a mind is the realization of a psychology in an agent. I decouple these terms from any connotation of human minds or actual physical realizability, admitting as "possible minds" agents including vending machines and logically omniscient intelligences. These conceptions are developed at length in Doyle (1982a)

The aim of rational psychology is understanding, just as in any other branch of mathematics. Where much of what is labelled "mathematical psychology" consists of microscopic mathematical problems arising in the non-mathematical prosecution of human psychology, or in the exposition of informal theories with invented symbols substituting for equally precise words, rational psychology seeks to understand the structure of psychological concepts and theories by means of the most fit mathematical concepts and strict proofs, by suspiciously analyzing the informally developed notions to reveal their essence and structure, to allow debate on their interpretation to be phrased precisely, with consequences of choices seen mathematically The aim is not simply to further informal psychology, but to understand it instead, not necessarily to solve problems as stated, but to see if they are proper problems at all by investigating their formulations.

This aim entails classifying sorts of agents and actions, classifying all possible minds, so that the detailed properties of an agent may be predicted from its fundamental classifications. Just as group theory seeks to classify the set of all groups in terms of their isomorphism classes and their relations to other mathematical structures, rational psychology seeks to classify the set of all possible minds and their relations to possible environments In either endeavor, a complete classification allows selection of standard representatives from each isomorphism class, representatives chosen to maximally facilitate their presentation and discussion. Put another way, rational psychology is one of the "sciences of the artificial," aiming to classify possibilities rather than to identify actualities. Classification can proceed without metaphysical doctrine, and as Courant and Robbins observe, some of the greatest achievements in physics have come as rewards for courageous adherence to the principle of eliminating superfluous metaphysics One must have a metaphysics, but it can be chosen, as well as inherited.

The method of rational psychology is to describe and study mental organizations and phenomena by the most fit mathematical concepts This does not mean pursuit of the mathematical tools for their own sake, nor forced application of pet mathematical abstractions, but simply the use of a precise language instead of vague formulations, and the borrowing of whatever analyses the current mathematics provides. The standards directing the investigation are those of mental importance rather than difficulty of proof or abstruseness or mathematical importance of the mathematical tools employed. If a result is not psychologically crucial, the difficulty of its proof does not lend it importance, and neither does the use of mathematical esoterica. But if analytic function theory captures the properties of some interesting agent more clearly than simple number theory, then it should not be shunned simply because of its relatively advanced position in mathematics curricula.

The method of rational psychology follows that of the mathematical study of mathematical concepts. One phrases subjects of investigation and specialized theories as sets of axioms about the constitution of agents These are called "constitutive assumptions" in modern rational mechanics. Rational psychology takes psychologies as givens for analysis, classification, prediction, and reformulation, rather than as mysterious qualities of agents to be discovered by experiment, computation, or philosophical speculation. These sets of constitutive assumptions can be formulated and studied for many external purposes: as ideals against which actual or constructed agents may be compared; as theories of actual or desired agents in special circumstances; as special aspects of actual or desired agents; and as approximations to the properties of actual or desired agents. Clean theories of special cases may "leave things out," but they so trade restricted range of applicability for enhanced accuracy within their domain of interest.

Comparison

Instead of giving a detailed sampling of important contributions to rational psychology, which would make this a long textbook rather than a brief notice, I list some of the areas I would include as contributions. Only a tiny fraction of this work has occurred within artificial intelligence, and rightly so, for artificial intelligence is only one of the newest of the fields of mental philosophy. Prominent among the areas with which the (ideal) student of rational psychology should be acquainted are (1) the sciences of rationality and rational agents, namely mathematical logic, metamathematics, and parts of mathematical economics (especially decision theory, game theory, utility theory, equilibrium theory, and social choice theory); (2) the sciences of mental representation and realizability, namely information theory, mathematical linguistics (both syntactical investigations and semantical studies), and the mathematical theory of computation; and (3) the sciences of mental ecology, for instance cybernetics and the new mathematical theories of perception. To these substantial theories, artificial intelligence contributes only a few smaller topics at present, such as the theory of perceptrons, search theory, and

theories of reasoned assumptions (see Doyle 1982b). These topics are still at the beginnings of their development and integration with other areas. As a non-example of rational psychology I offer the theory of measurement. This theory appears prominently in texts on mathematical psychology, but is really no more relevant to psychology than to physics or demography It supplies analysis of methodological questions and experimental procedure, but has little bearing on the nature of mental or physical entities. This does not reflect badly on the theory of measurement, any more than the irrelevance of ceramics to psychology reflects badly on ceramics.

I build on this non-example of rational psychology to make the principal aims and methods of rational psychology clearer by contrasting them with the principal aims and methods of related fields These brief characterizations are all somewhat unjust, for fields are populated by people with mixed interests; but they serve nevertheless to illustrate different emphases To begin: the modern discipline of Psychology is the experimental investigation of human psychologies, with studies of other animals as paths to humans. Humans and experiment form the focus of Psychology, rather than all possible minds and mathematical analysis The philosophy of mind, while employing conceptual (but typically not mathematical) analysis, also focusses on humans almost exclusively In economics, where mathematical analysis has become standard, the focus is on rational agents, individual and collective, rather than on agents in general. Similarly, logic and metamathematics look to rationality, not general psychologies Chomskyan linguistics is explicitly oriented toward the human mind, via the mechanism of language The neurosciences are similarly both human- and mechanism-oriented Cognitive science, to the extent that it admits a consensus, is an amalgamation of the human-oriented fields and artificial intelligence Artificial intelligence itself, which from its name might seem the natural companion to the aims of rational psychology, is quite fragmented in aims, but almost universally oriented toward recursive realizability of agents in modern digital computers Its subfield of cognitive simulation is explicitly human-oriented, and its subfields of formal reasoning, automated deduction, and "theorem proving" are all oriented towards issues of rationality rather than psychologies in general. "Reasoning" means deduction to almost all involved The focus of the field on gaining insight from computational experience is valuable, for exact analysis always has current limits, but few pursue any exact analysis at all

Benefits

Rational psychology offers a number of practical benefits The first of these is that of formal, precise statements of artificial intelligence problems, theories, and techniques. Formal specifications of program intent and proofs of program correctness are well-known in computer science These concepts, though hardly a panacea, now allow concise and correct description of systems whose understanding previously required apprenticeship and experience These exact formulations permit variations in problem and solution to be studied as technical questions rather than as banners in battles between methodologies and world-views. Mathematical formulation of concepts has hardly been prominent in artificial intelligence, with good reason For the most part, complete ignorance prevails about the appropriate mathematical structures to employ in formulating psychological notions, and there is every reason to suspect that many new mathematical notions must yet be invented in order to develop current informal psychological theories in precise terms To draw a parallel, no matter how much one hoped to assign meanings to computer programs and their components, all early attempts to do so foundered on the reflexive nature of the domain of all computable functions, so that every proposal prior to Scott's discovery of appropriate models was either obviously inadequate or of such complexity as to be of doubtful correctness. Unfortunately, for most of artificial intelligence, suitable mathematical tools are similarly undiscovered, so no matter what their standards when discussing computer science, many researchers find that doing artificial intelligence requires abandoning the usual crutches of confidence for wild and woolly adventures in intellectual hinterlands Some never return to tell their tales, and some return speaking in tongues to the rue and mutterings of the stick-at-homes. Formal specifications may not be an immediate path to benefits, for discovery of the appropriate concepts doubtless requires much toil. But someday, it must be done.

The second benefit rational psychology offers, even to the hard-core hacker, is savings in time and resources. Mathematics can be viewed as the science of avoiding unnecessary calculation, and rational psychology can be used as a way of avoiding some labors of programming and computation. It is commonplace in artificial intelligence research that systems are developed at costs of man-years and CPU-months, and when finished, their authors discover trivial examples of fundamental inadequacies and seemingly unmotivated limitations of abilities that to remedy would require the effort all over again. One cannot hope to discover all difficulties with a pet idea through thought alone, nor hope to avoid all unconscious intellectual blinders, but cultural practice in artificial intelligence calls for implementing ideas as sufficient means to "understanding" them. Often some inadequacies and tacit limitations come to light in this process, but diluted by months or years of wondering where the next CONS is coming from. Consider instead a cultural imperative which called for three weeks of pure critical (even adversary) thought and strict abstinence from computers prior to beginning any important implementation effort The problems of artificial intelligence would not become any easier, but progress might be faster, since one might trade a week of analysis for a year of wasted programming. Socrates might well have said "The unexamined idea is not worth pro-

gramming," and had the Athenians personal computers with LISP-controlled graphics they might well have sentenced him anyway There is great contrast between the pleasures of programming and the tedium of analysis, between the challenge of the mysterious bug and the death of a beautiful hypothesis at the hands of an ugly fact

Rational psychology also offers improved communications The frequency of reinvention of ideas in artificial intelligence is legendary. While it is unreasonable to expect (and undesirable to attempt) to make reinventions rare occurrences, artificial intelligence clearly seems extravagant It is not alone in this. There is the old joke in computer science about the result that was lost because it was only published four times But even the magnitude of the problem is unclear Not only do researchers lack deep understanding of their own proposals, but they usually cannot understand those of others either. This incomprehension is not due to stupidity, but to the vague, metaphorical terms on which the field relies in the absence of precise, formal vocabularies for presenting theories. In mathematics, physics, and many other sciences, papers, if properly written, define concepts in terms of the accepted vocabulary, state claims or discoveries, and then leave comprehension up to the intelligence and motivation of the reader. In artificial intelligence, even conscientiously written papers can be unintelligible no matter how capable and motivated the reader, for much of the accepted vocabulary is about as precise as that of poetry, and about as substantive as that of advertising copy If we had adequate mathematical concepts, if we had conventions for clear, exact statements of problems — two large ifs — then we could hope for reduced reinvention, more rapid communication, comparison, and reproduction of ideas, and a true chance to build on the work of others: things all taken for granted in other fields

Conclusion

A mathematical, analytical enterprise like rational psychology is not for everyone Indeed, rational psychology feeds on intuitions gained only through experience, so it makes no more sense for everyone to abandon the usual efforts of artificial intelligence and cognitive science than for all physicists to forsake experiment and experience in favor of rational mechanics On the other hand, rational psychology need not be purely parasitic, for its pursuit may someday advance the construction of thinking machines, much as aerodynamics has advanced the construction of flying machines But these practical benefits cannot be realized without effort At least some people must stray from the usual investigations of artificial intelligence and cognitive science, and their work must be judged by the aims and methods of rational psychology instead of by those of artificial intelligence and cognitive science. I would not bother to invent the label "rational psychology" for these aims and methods, except that they *are* somewhat different from the usual ones of artificial intelligence and cognitive science, and more easily understood and encouraged when explicitly recognized For example, questions about implementation status or experimental verification of theories are legitimate questions for artificial intelligence and cognitive science, but not for rational psychology, even though the same theories may be under discussion As with chemistry and cookery, mere recipes for constructing machines and men do not guarantee understanding the product And for rational psychology, the main question is whether the theories have been adequately understood.

References

Courant, R and Robbins, H , (1944) *What is mathematics? An elementary approach to ideas and methods*, London: Oxford University Press

Doyle, J. (1982a) The foundations of psychology, Pittsburgh: Department of Computer Science, Carnegie-Mellon University

Doyle, J (1982b) Some theories of reasoned assumptions: an essay in rational psychology, Pittsburgh: Department of Computer Science, Carnegie-Mellon University

Doyle, J (1983) What is rational psychology? Toward a modern mental philosophy, Pittsburgh: Department of Computer Science, Carnegie-Mellon University

Hartmanis, J (1981) Remarks in "Quo Vadimus: computer science in a decade," in J F Traub (ed), *Communications of the ACM* 24, 351-369

Minsky, M. (1962) Problems of formulation for artificial intelligence, *Proc Symp on Mathematical Problems in Biology*, Providence: American Mathematical Society, 35-46

Nilsson, N. J. (1980) The interplay between experimental and theoretical methods in artificial intelligence, Menlo Park: SRI International, TN 229

Truesdell, C. (1958) Recent advances in rational mechanics, *Science* 127, 729-739

(Continued from page 49)

Second National Conference on Artificial Intelligence, Pittsburg, PA., 300–331

Novak, G S (1982) The GEV Display Inspector/Editor Tech Rept. HPP-82-32, Heuristic Programming Project, Computer Science Dept , Stanford University

Novak, G. S (1983) *Knowledge-Based Programming in GLISP* American Association for Artificial Intelligence, *Proc Third National Conference on Artificial Intelligence*, Washington, D C , in press

Teitelman, W (1978) *INTERLISP Reference Manual* Xerox Palo Alto Research Center

Wulf, W. A., London, R , and Shaw, M. (1976) An Introduction to the Construction and Verification of Alphard Programs *IEEE Transactions on Software Engineering* SE-2, 4