Conceptual Parallels Between Philosophy of Science and Cognitive Science: Artificial Intelligence, Human Intuition, and Rationality

Joseph McClure

Research in the cognitive sciences is founded on the assumption that the brain is, essentially, a very complex kind of computer—an information processing organ that instantiates abstract computational operations over formal symbolic structures. This assumption is closely tied to a theory in both philosophy of mind and cognitive science known as the *computational theory of mind* and has guided the majority of empirical research in artificial intelligence and other areas of cognitive science.

My goal in this paper is to illustrate the conceptual, analogic relation between the notion of confirmation and theory change, as debated in the philosophy of science, and the symbolic computational theory of human intelligence, as debated in cognitive science and the philosophy of mind. My claim is that these two issues have a remarkably similar conceptual structure—indeed, many of the same (or similar) phrases and characterizations can be found in the literature on both topics—and that an understanding of this deep similarity can help illuminate an underlying flaw in the conceptual cores of both issues, as debated in their respective fields of study. Specifically, I will examine the attempt of philosophers of science to formulate and rely on a purely syntactic, symbolic "language of science" to understand the rationality of theory choice and confirmation. I will argue that this mirrors the attempt of cognitive scientists and artificial

Joseph McClure graduated with a degree in philosophy and a degree in biopsychology, cognition, and neuroscience from the University of Michigan. His primary philosophical interests include philosophy of mind and philosophy of cognitive science. He is currently applying to Ph.D. programs in philosophy.

intelligence researchers to formulate an algorithmic, mechanistic model of human intelligence as the formal operations of a symbol-processor. As a consequence of this parallel conceptual structure, each attempt is vulnerable to important criticisms leveled at the other.

In this essay, I will attempt to reveal the conceptual similarities between both issues. In order to do so, I will begin with cognitive science and philosophy of mind by discussing the historical and philosophical foundations of research in artificial intelligence and the symbol-manipulation view of the mind. I will then present one of the most significant critiques of this view, which deals with the inability of traditional rule-based AI to account for human expertise and intuition. Following this discussion, I will move to philosophy of science by providing an overview of some issues surrounding the rationality of theory choice, focusing in particular on the perceived conflict between intuition and rationality. The remainder of the paper will consist of a discussion of the conceptual parallels between this issue and the traditionally held assumptions behind research in artificial intelligence. I will conclude by proposing some methodological and philosophical implications of this parallel, the most significant of which deals with our appreciation of the non-rational aspects of human intelligence.

I. Foundations of Artificial Intelligence Research

In this section, I'll briefly review the origins of the research program known as artificial intelligence (AI) and its relationship with symbol-manipulation theories of mind. The theoretical basis of the artificial intelligence and cognitive simulation research programs is the notion of computation, which can be traced as far back as Descartes and has been articulated in various forms by Pascal and Leibniz, among others (Dreyfus 53). The basic notion of computation as the rule-governed operations of a system composed of simple elements is a powerful idea. With the advent of digital computers around 1950, logicians, psychologists, and computer scientists like Alan Turing, Allen Newell, and Herbert Simon finally had an appropriate tool with which to empirically test the power of their symbol-manipulation theories (53). Since the symbols in computational systems are manipulated solely on the basis of their syntactic features—their formal properties—they can be made to represent just about anything, including objects and relationships found in the world (as long as the objects can be represented symbolically and the relationships can be specified as formal rules).

The representational utility and processing power of computers, along with an awareness of the systematicity and formal structure of certain human mental capacities (most importantly language), led both scientists and philosophers to compare the abilities of the mind with the abilities of computational systems. The culmination of early research in cognitive systems modeling, combined with a mature and sophisticated understanding of the principles of formal computational systems, led to the development of the computational theory of mind (CTM) in philosophy and the empirical and theoretical research program of artificial intelligence.

CTM asserts that all mental processes, unconscious or conscious, are computational processes, albeit of a highly complex and sophisticated variety. The goal of psychological and artificial intelligence research, then, is to figure out which rules and algorithmic procedures are responsible for mental capacities and abilities such as planning, choosing goals, and formulating strategies. Cognitive scientific research, including research in AI, operates largely under the "information processing assumption" that conclusions drawn from sets of rules and lists of features are at the heart of intelligent behavior and mentation (27). Despite the failures of AI to live up to its own expectations, the majority of cognitive scientific research has retained this core conceptual framework—the view of the mind and intelligent behavior as the rule-governed operations of a formal symbol-manipulating system.

II. Criticism of the Research Program in Rule-Based AI:
Human Expertise and Intuition

Although the physical symbol system hypothesis still appears to be the guiding paradigm in cognitive science and the majority of empirically-informed philosophy of mind, it faces a number of important and underappreciated challenges. Among these challenges, the most notable are offered by the philosophers Hubert Dreyfus and John Searle. Dreyfus's critique, presented in *Mind Over Machine* and co-authored with his brother Stuart, details the methodological and philosophical failures that have plagued research in traditional rule-based AI as a model of human intelligence and expertise. Searle's attack on the computational, symbol-manipulation view of the mind is no less significant, but the parallels I am interested in drawing are better explained in Dreyfus's terms.

According to Dreyfus, the physical symbol-systems view of the mind does not account for our ability to act as experts: our capacity to engage

with situations in an intuitive, highly fluid manner characterized by intuitive appraisal and action without rational deliberation. Central to this claim is his account of skilled behavior as a progression from novice to expert. For my purposes, we will focus on the "competency" and "proficiency" stages, but I will briefly characterize the overall progression. The first stage, "novice," is the beginning of a person's experience with the skill to be learned. The example that Dreyfus frequently uses is that of chess playing (Dreyfus 22). The novice chess player learns the basic rules of chess, and their application of those rules is deliberate and conscious. The rules are context-free; that is, they do not make reference to the aspects of any particular situation, but they are general rules that give a basic structure to the problem domain. As the person progresses in learning a skill, they gradually progress from the detached, analytic perspective of the novice (stage 1) to the involved, intuitive, situated perspective of the expert (stage 5). Intuition, according to the Dreyfus brothers, "*is neither wild guessing nor supernatural inspiration, but the sort of ability we all use all the time as we go about our everyday tasks*, an ability that our tradition has adjudged inferior to masculine rationality" (29, italics in original).

Stages 3 and 4 are the two that are most relevant for my purpose in this paper. At stage 3, the "competent" performer still makes use of the kinds of "context-free" facts that the novice so laboriously utilizes, but is now able to organize these facts, which are much more numerous than before, into coherent, goal-directed plans (65). Certain situational elements are thereby made salient, but only through the deliberate application of rules that are themselves context-free.

At stage 4, "proficiency," the performer has achieved the ability, through extensive experience and association with past situations and responses, to engage with the task or problem-space in an intuitive, deeply involved capacity. Features of the current situation present themselves as significant based on the holistic appraisal and situated recognition of the performer, not through any calculative procedure, but simply because of associations between the present situation and similar situations in the past, including things like which plans worked, which didn't, and what came next (28). The proficient performer has not abandoned rational intelligence, but is usually so involved in the task that there is no need for deliberate reflection and analysis (which is not to say that the performer is then acting *ir*rationally— *a*rational might be an appropriate term). The proposed analysis of skilled behavior as a progression from detached, analytic reflection to involved, intuitive action is in direct conflict with the information processing model of human intelligence assumed by proponents of traditional AI. According to this information processing model, improvements in performance are

the result of the accumulation, refinement, and increased processing speed of rules similar to those that the beginning performer follows consciously (65). My own view is in line with the Dreyfus brothers, who believe "that a move to the new level of performance coincides with a shift from the logical processing of atomic facts to the direct holistic association with the current situation of what experience has shown to be an appropriate response" (66).

III. The Issue of Theory Choice in Philosophy of Science

In this section, I will review Hempel's attempt to develop a logical, deductive theory of the criteria of confirmation of scientific hypotheses. I will contrast this logical empiricist stance with Kuhn's historical analysis of theory choice and his notion of the "paradigm."

A. Logical Empiricist Methodology: Hempel on Confirmation

Hempel's "Studies in the Logic of Confirmation," is an exemplary illustration of the conceptual framework of logical empiricism, in particular the significance accorded to rational, deductive methodological principles governing scientific explanation and theoretical structure. According to Hempel, the scientific enterprise is characterized by the formation and testing of empirical hypotheses—using the methods of logic and deduction to rationally and objectively determine the laws and principles of nature that provide the structure of scientific theories (35, 55-6). In the paper, he presents an extended analysis of the notion of confirmation, i.e., what it means for a hypothesis to be supported or validated. The result of his analysis is a logical, analytic description of the criteria involved in the notion of confirmation in the methods of science. His goals are characteristic of the logical empiricist view that rationality and objectivity are the most important and highly valued principles of the scientific method and that subjective standards, such as a "sense of evidence," should dismissed (37). What I find more striking, and more important for the purposes of this paper, is his subsequent claim:

> In fact, it seems reasonable to require that the criteria of empirical confirmation, besides being objective in character, should contain no reference to the specific subject matter of the hypothesis or of the evidence in question, it ought to be possible, one feels, to set up purely formal criteria of confirmation in a matter similar to that in which deductive logic provides purely formal criteria for the validity of deductive inferences. (37)

Hempel's proposal here is that the criteria by which scientific hypotheses are confirmed should be conceived of as analogous to the criteria used in deductive logical inference—criteria that are powerful precisely because of the formal, abstract character of logical systems. By drawing on the tools of logical analysis, Hempel attempts to give a description of the system of criteria needed for empirical, scientific confirmation, structured so as to constitute a system that preserves above all others the values of rationality, objectivity, and generality. These characteristics are all derived from the formal, syntactic structure of the methodological system. As such, we can ignore all aspects of the problem that are extraneous to the logical, syntactic structure of the description (44). I will now turn to Kuhn's views on the methodology of science.

B. Kuhn: Paradigms and Theory Choice

Kuhn's *The Structure of Scientific Revolutions* constitutes an attempt at reconciling the methodological theories proposed by philosophers of science with the actual history of scientific theory and research. This results in an appreciably different view of the nature of science than that proposed by logical empiricists like Hempel. For this paper, I will focus on Kuhn's conceptualization of the scientist in crisis. During "normal science," when the reigning paradigm provides a clear theoretical structure, the scientist is engaged in the activities that we typically think of when we think about scientific research. These are essentially problem-solving enterprises, where the problems are identified and investigated in the context of the guiding paradigm. The paradigm gives structure and significance to the "world of the scientist" by providing a theoretical and conceptual framework. Kuhn writes:

> As a result of the paradigm embodied experience of the race, the culture, and, finally, the profession, the world of the scientist has come to be populated with planets and pendulums, condensers and compound ores, and other such bodies besides. Compared with these objects of perception, both meter stick readings and retinal imprints are elaborate constructs to which experience has direct access only when the scientist for the special purposes of his research, arranges that one or the other should do so. (176)

Under a paradigm, the methods of science are perhaps something like the formal, objective inferences from sets of data to general hypotheses described by Hempel and other strict methodological empiricists. But when the paradigm changes, there is a period in which the scientist no longer has the comfortable, given structure of a solid theoretical framework.

For a new paradigm to arise, something like an intuitive conceptual leap must be made—a process that cannot consist solely in the reinterpretation of existing data (169). Kuhn's idea, based on historical analyses of periods where scientific theories underwent large-scale transformations, is that new paradigms are born primarily from the kind of intuitive insight described by Dreyfus. Kuhn argues that while this paradigm-generating insight is based on experience obtained within the context of the old paradigm, the new paradigm is generated when the scientist is able to intuitively "gather up large portions of that experience and transform them to the rather different bundle of experience that will thereafter be linked piecemeal to the new paradigm but not to the old" (171). Like Dreyfus, Kuhn emphasizes the importance of intuition in human cognition and expertise.

IV. Synthesis: Conceptual Parallels Between Issues in Philosophy of Science and Cognitive Science

My central claim in this paper is that there are interesting and illuminating parallels between the conceptual and methodological issues present in cognitive science (particularly AI research) and those in philosophy of science, and that these parallels reveal a shared weakness. Now that we have a sense of these issues, I will explore how two of the most significant parallels illuminate this common, underlying conceptual flaw.

Parallel I:
Symbol-Systems Model of Intelligence || Logical Empiricsm

It is perhaps already apparent that the theoretical and methodological stances assumed by traditional AI researchers and by logical empiricists share an almost dogmatic emphasis on the importance of rationality and formal structure. Within cognitive science and artificial intelligence research, the dominant theory of mind equates all mentation with computation—in particular, the computational operations of a formal system over purely syntactic representations. These computations, which are merely mental processes, are asserted to be abstract, objective descriptions of definite, specifiable rule-governed algorithms that are instantiated in the neural substrate of the brain. Similarly, within the philosophy of science, logical empiricism asserts that the methods of science can be characterized as sets of logical relations and rules for judging the acceptability of observation reports of empirical data.

Proponents of the symbol-systems theory and logical empiricism both seek to construct objective, rule-governed models of the activities of apparently structured systems: just as cognitive science is concerned with the system of the mind, logical empiricism is concerned with the system of hypotheses and practices that make up the scientific enterprise. My claim is that both research programs share a crucial and perhaps misguided assumption: namely, that the rationally objective, rule-governed, formally structured elements of a system constitute a conceptual framework capable of providing a complete understanding of the system as a whole. The problem with this view is that it assumes the formal aspect of any system, abstracted and decontextualized, gives a necessary and sufficient description of the system. This precludes an understanding and appreciation of the non-formal aspects of a system, whether the system is the mind/brain or scientific methodology. This is a mistake, as Dreyfus makes clear, because it is readily apparent that a significant part of our expert intellectual activity *cannot* be adequately analyzed in formal, symbolic terms: "Human beings," he writes, "exhibit a flexibility, judgment, and intuition that resist decomposition into specification and inference and have proved equally difficult to instill into logic machines" (Dreyfus 63).

The pursuit of a strictly rule-governed methodology of science, like the pursuit of a strictly rule-governed model of human intelligence, is often construed as an exercise in, and demonstration of, the analytic power of objectivity and rationality. This is especially the case where such power is applied to the goal of constructing a self-contained, extensive system of rules. This objective stems from a strict adherence to the ancient philosophical tradition that upholds calculated rationality as the finest aspect of human intellect. As such, attempts to construct rule-based computational systems that match human level expertise have run into serious problems. However, these problems have been largely pushed aside with the assumption that more sophisticated and powerful systems will solve them (Dreyfus 10).

Nevertheless, it is precisely those aspects of our intelligence that resist rational, rule-based description that characterize a substantial portion of our capacity to respond skillfully to changing situations, exemplified in expertise and intuitive reaction. Simply put, we must recognize that "Hunches and intuitions, and even systematic illusions, are the very core of expert decision-making," and that these peculiarly human abilities are often difficult to capture successfully using computerized models and algorithms (Dreyfus 10).

The conceptual tension displayed here in the cognitive sciences is mirrored almost perfectly in philosophy of science and the logical

empiricist movement. Attempting to describe or prescribe scientific methodologies in purely formal terms, abstracted from context and expressed in a logical "language of science," is as misguided as attempting to model human intelligence in terms of formal algorithms and rule-governed manipulations. Both attempts neglect the embodied, involved nature of human intellectual activity—our ability to intuitively discriminate the significant aspects of the situations we find ourselves in and respond in a goal-directed, but fundamentally meaningful manner. This is far from the detached, calculative rationality and inference from objective facts to formal hypotheses by the rules of logic that is assumed in both traditional AI and logical empiricism to be all that there is.

Parallel II:
Human Intuition and Skilled Coping || The Kuhnian Scientist

The preceding discussion focused on the parallel between the shared adherence to formal, objective analyses in philosophy of science and in AI research. Here, I want to address the corresponding parallel between Dreyfus' critique of the physical symbol systems hypothesis in AI and Kuhn's critique of the characterization of the scientific method as a purely logical, analytic enterprise. The conceptual parallel here deals with Dreyfus' stages of skill acquisition and Kuhn's description of the behavior of the scientist. Hempel dismisses any non-rational consideration in the scientific process as neither necessary nor sufficient for confirming the validity of a hypothesis, because such considerations are based on one's personal, subjective "feelings of conviction" which are by no means guaranteed to remain constant over time, and which often vary from person to person (Hempel 37). I think that this dismissal is too quick by far.

In any well-structured domain, a suitably programmed computer can instantly achieve the level of novice and even the level of competence if given enough data and rules. Just so, a rule-based methodological theory that assumes an objective realm of experiential data and a logical inferential and confirmatory procedure might be quite suitable to areas of science that are well-structured and lend themselves to formalization in this manner. Having drawn this parallel, however, I claim that the methodology of scientific confirmation and theory choice during periods of scientific revolution and paradigm change cannot be characterized in the purely formal manner desired by logical empiricists like Hempel.

To illustrate the point above, I wish to draw on Dreyfus' theory of skill and expertise and present a characterization of the scientist in the

midst of a shifting paradigm. During this transition, the scientist's world of experience shifts radically from the structural and theoretical cohesiveness given by the paradigm before the time of crisis. Therefore, he or she can no longer rely on detached, calculative rationality in the search for formal rules that relate objective experience to general hypotheses—this is the scientist at approximately the stage of the "competent" performer. The discovery of a novel paradigm relies instead on the intuitive value judgments and involved appraisal at the stage of (at least) the "proficient" performer. Nothing less than the situational awareness that comes with vast experience—*involved, embodied, reactive experience*—can provide the sense of significance and salience that is necessary for the ability to cope with the sudden uncertainty of formal structure that characterizes scientific revolution.

V. Conclusions and Implications

Both of the parallels explicated above deal, broadly speaking, with the perceived conflict between rationality and intuition. The scale has weighed heavily in favor of rationality as the human animal's most prized trait; we see abundant evidence of this favoritism in the cognitive sciences, where research in artificial intelligence is still early in the process of according intuitive, embodied intelligence the status it deserves—a place beside, not below, our calculative rationality. In philosophy of science, the insistence on detached, rule-based methodology and theory is perhaps better placed; but the scientist is human, and the philosopher of science must give due consideration to the involved, holistic understanding that defines human expertise and intuition. Otherwise, we limit our understanding to just those formal aspects that are assumed to be all that matter, while the substantive, non-representational background that structures and grounds our abstractions is left out of the picture. For philosophers of science attempting to understand and analyze the methods of scientific practice, the implications are profound. We must acknowledge the validity of expert wisdom, even if this intuitive understanding cannot always be expressed in a form suitable for objective scrutiny. Failing to give this subjective expertise the respect it deserves means that we are missing out on an important and uniquely human intellectual tool. If we don't recognize this capacity's pervasive impact on philosophy, scientific research, and everyday life, our understanding of these fields will remain incomplete at best.

Works Cited

Dreyfus, Hubert. L. and Stuart E. Dreyfus. *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York: Macmillan, 1986.

Hempel, Carl G. "Studies in the Logic of Confirmation." *Philosophy of Science: An Anthology*. Ed. Marc Lange. Malden, MA: Blackwell, 2007. 34–61.

Kuhn, Thomas S. *The Structure of Scientific Revolutions*. Chicago: U of Chicago P, 1970.