Review of “Having Thought”

Zsolt Bátori  
Rutgers University

Follow this and additional works at: https://commons.pacificu.edu/eip
Part of the Philosophy Commons

Recommended Citation
In *Having Thought* John Haugeland discusses a variety of problems concerning understanding and intelligibility. The book consists of thirteen essays, of which eleven have been previously published. The essays are arranged more or less chronologically (ranging from a previously unpublished paper that was written in 1977 to the last and most recent work that is new in the volume), and they are also organized into groups, under four heads: Mind, Matter, Meaning, and Truth.

In the first three essays Haugeland examines the nature of the human mind as an understander and the problems cognitive science must face when trying to account for the human mind in terms of an information processing system. The human self is argued to be existentially holistic; what makes it possible to interpret human behavior and natural languages intentionally is that humans have and care about their whole life. Intentional interpretation of natural languages is thereby distinguished from the interpretation of information processing systems such as computers. As Haugeland puts it: "the trouble with artificial intelligence is that computers don’t give a damn" (p. 60).

Haugeland investigates the prospects of cognitive science from two points of view. First he examines the nature of the specific kind of scientific inquiry that attempts to explain the mind in terms of an information processing system. The kind of explanation specific to cognitive science is argued to be systematic (and at some level morphological) that is analogous to the explanation of the functioning of an engine, for instance. This kind of investigation does not necessarily involve the familiar derivational-nomological explanations of science (based on presupposing regularities as laws that apply in a variety of situations which exhibit some specific features). Nevertheless, explanation in cognitive science is still a perfectly legitimate form of scientific explanation. It proceeds by systematic reduction, that is, by reducing the mind to functional components (subsystems) and by providing morphological explanations of the "bottom" components of the system. Attributing meaningfulness to information processing systems (intentional interpretation) is then argued to be justified if the behavior of the system can be interpreted as part of an overall pattern that makes sense. Since intentional interpretation is testable, it is an empirically legitimate form of scientific explanation.

Although the mind can be understood and accounted for in a scientifically legitimate way, Haugeland maintains that cognitive science may still be unable to provide us with a sufficient explanation for a number of features of the mental. There are mental states (moods, for instance) that cannot be interpreted as linguistic or quasi-linguistic. Although moods affect cognitive states, they themselves are not cognitive states, and cognitive science may not have the adequate theoretical tools to handle them. Similar worries
might be raised about skills (playing an instrument, typing, etc.) whose development requires practice, examples and talent (which is to be distinguished from intelligence), while no articulate theory is necessary (or even developed in some cases) in order to acquire them.

Haugeland also argues that the trouble with artificial intelligence is that it is not clear how an information processing system can be an understander. Intentional interpretation is argued to be holistic; for instance, when we say that a chess playing device "tries" to make good moves, "wants" to win, etc., such attributions are made in the context of the game. Common sense holism of the interpretation of natural language sentences is based on interconnected and cross-referenced structures; concepts understood as encyclopedia entries. Situation holism requires us to take into account an indefinite number of possible situational modalities in order to disambiguate complex narrative structures. Finally, people’s acts can only be understood to be significant and meaningful if their acts are interpreted as events "in the biography of a whole, historical, individual", since "[o]nly a being that cares about who it is, as some sort of enduring whole, can care about guilt or folly, self-respect or achievement, life or death" (pp. 58-59). This is what Haugeland calls "existential holism" which is necessary for understanding stories. Computers (understood as information processing systems), however, "don’t care", and Haugeland argues that common sense holism, situational holism and existential holism present increasingly difficult steps for AI.

The second group of essays is centered around two problems. First, Haugeland argues against the thesis that analog systems (such as brains or other physiological systems) can be digitally simulated to any degree of precision. Second, Haugeland considers and rejects the token identity theory and argues for a weaker version of supervenience-based physicalism.

In order to consider the question of simulating analog systems digitally Haugeland first provides a careful analysis of the nature of the two kinds of systems. Digital devices, such as alphabets, musical notations, electrical switches, arabic numerals, and the like, are argued to exhibit three characteristics: (1) perfect copyability, (2) complexity (being complexes of standard components), and (3) medium independence (equivalent structures can be instantiated in different media). Analog devices, such as photographs, scale models, loudspeakers, slide rules, etc., also have three common features: (1) smoothness or continuity of variations, (2) sensitivity for every difference within the relevant range, and (3) dimensionality (certain dimensions of variations, such as chemicals that do not affect tonal distribution, are not relevant).

Haugeland then defines both a digital and an analog device as a set of types. On the basis of the above distinctions, he argues that the writing and reading of tokens of these types are positive and reliable procedures in the case of digital devices; the procedures either succeed or fail. That is, success or failure is unqualified, approximations (success to some degree, for instance) are not allowed. In the case of analog devices, however, writing and reading are approximation procedures, not positive ones. Approximation procedures always have some degree of deviation from perfect success, that is, the margin of error can never be zero.

Physiological systems (metabolic systems, brains, etc.) are complexes of millions or billions of balances, interactions and reactions among their molecular components. Since the digital simulation of even one organic molecule is an overwhelming task, Haugeland argues that digitally simulating the organic molecules of a human brain (or whole body) itself would be implausible, not to mention the combinatorially growing number of computations that would be required to simulate the interactions among the molecules. Therefore, Haugeland concludes, it is mistaken to hold that analog systems can be
The second main issue Haugeland addresses in this section of the book is the token identity theory of supervenience-based physicalism. The token identity theory holds that each mental event token is numerically identical to some physical event token, and mental and physical event predicates range over the same domain, nevertheless, "the ‘kinds’ or ‘types’ intended by mental event predicates are not nomologically or necessarily equivalent to any kinds or types intended by physical event predicates (no matter how complex)" (p. 91).

In order to show that "weak supervenience" (that is, physicalism without identity theory) is a feasible and coherent position Haugeland provides a model (by describing possible worlds and two languages relative to them) in which supervenience holds (one language supervenes on the other) but the token identity thesis is false. After having an alternative theory in hand, Haugeland offers a number of arguments against the token identity theory. For instance, he considers distinct waves hitting a single cork from two opposite ends of a swimming pool. Although the two waves hitting the same cork at the same time can be described as two distinct robust events (with different causal histories), particular positions or velocities at the molecular level are not constituents of one or the other wave hit. Haugeland therefore concludes that "the individuals or tokens of which our sentences are true are just as ‘relative’ to the level of description as are the kinds or types into which those sentences sort them" (pp. 102-103).

The overall theme of the third group of essays is the nature of the connection between the mind and the world. Haugeland provides an overview of the accounts of original (non-derivative) intentionality as well as a number of considerations about representations and meaning. This section of the book includes an interesting chapter on logical, pictorial and distributed representations, and on the consequences (in philosophy of mind and in cognitive science) of accounting for them by exploring the characteristic structures of their represented contents, instead of investigating the relation between representations and their contents. However, the main issue in this section of the book is the locus of intentionality and intelligence.

The overview Haugeland gives divides the field the following way. What needs to be explained is how it can be that some things are "of" or "about" others. In other words, materialism requires an account of representation, since intentionality seems to be incompatible with the view that the world consists of meaningless physical particles. The intentionality of individual states or events cannot be explained independently, therefore, theories of intentionality must identify the nonaccidental larger pattern in which the intentionality of individual states or events fits. Haugeland describes three ways of accounting for that larger pattern. Neo-Cartesians argue that mental states have their content in virtue of their systematic relation to one another, and that the overall pattern bears some relation to the world. Neo-behaviorists identify the nonaccidental systematic larger pattern in the interactions between the intentional system and its environment. Finally, neo-pragmatists locate original intentionality in the culture or way of life (including social practices, institutions, artefacts, etc.) of a community.

Haugeland argues that the Cartesian division between mind and body as well as mind and world may well be the source of more confusion than insight with regard to the nature of intelligence. The reason for this, so Haugeland suggests, is that it is possible that the division (including the notion of mind/world and mind/body "interfaces") is just as arbitrary as it would be to divide an electronic device, such as a television set, into equal slices along each axis, and then to hope that the examination of this division and of the resulting "parts" would provide some insight into its functioning. Another example would be to
divide organizations (governments, universities, corporations, etc.) on the basis of some corporeal boundaries, such as buildings, when (especially today, when communication networks are integral parts of such organizations) the physical location of people and data are often irrelevant, and therefore a division or department of an organization may well be (and often is) spread across a number of buildings, states, or even continents. Arbitrary division on the basis of corporeal boundaries would, in fact, prevent the understanding of the functioning of the organization in question. Haugeland not only questions the Cartesian division between the mind and the body, but also argues (along the lines of neo-pragmatism) that human intelligence and competencies are only intelligible in terms of their higher unity, where the higher unity is not to be understood as merely the sum of individual knowledge and competencies. The shapes and strengths of tools, the practices for building and using them, the structure of institutions (the relations among their parts, the practices established by institutions, etc.) are argued to be integral parts of human intelligence. Haugeland concludes that intelligence is externalized in ways that cannot be accommodated in the Cartesian framework. "Mind, therefore, is not incidentally but intimately embodied and intimately embedded in its world." (P. 237.)

The last section of the book includes Haugeland's recent work on understanding and on how understanding can be objective. The discussion of the broader question of the nature of human understanding involves a number of specific issues, such as a novel classification of different kinds of intentionality, rule following, and the role of subjectivity in objective understanding.

In the course of exploring both agreements and differences among the views of Dennett, Searle, and himself, Haugeland develops a theory of human understanding that is centered around the "commitment" of the understander. "Understanding a domain and its entities is understanding the principles according to which that domain and those entities are constituted; and such understanding can be nothing other than a commitment to those principles. Intentionality presupposes a committed stance because intentionality -- meaning -- presupposes understanding." (P. 301.) For instance, participating in (understanding of) a game of chess involves being committed to constitutive standards (in this case to the constitutive rules of the game). Commitment to standards is argued to be the foundation and essence of intrinsic intentionality, that is, of the kind of intentionality that is specific to humans.

Haugeland agrees with Dennett and Searle that there are at least three different ways we can talk about intentionality. Intrinsic intentionality is what humans (and perhaps higher animals) share, derived intentionality is what public symbols (and, according to Searle, computers) have, and as-if intentionality is the metaphorical use of the term (e.g., when we say that the saplings are "trying" to grow up to the sunlight). Haugeland, however, suggests that there are more distinctions to be made, and that animals and computers are to be reclassified as having "ersatz" intentionality. On the one hand, Haugeland argues that even though a robot's resources, goals and standards are preprogrammed, it can produce and use representations. Manipulating and producing representations separates it from public inscriptions, so its intentionality is more than derived. However, the standards in virtue of which its states can be understood as intentional and normative are external to it (given to it by the programmer), therefore its intentionality cannot be considered intrinsic either. Ersatz intentionality is the kind between derived and intrinsic. On the other hand, the intentionality of animals is understood relative to standards that we establish for them; "animals do not commit to constitutive standards, hence do not submit themselves to norms, and do not understand anything". (P. 303). Therefore, so Haugeland concludes, the intentionality of animals cannot be considered intrinsic, but only ersatz.

Haugeland also argues that following rules (in order to play games, for instance) and conducting
(objective) empirical science require a specific kind of first-person (subjective) involvement and commitment. Existential constitutive commitment is *sui generis* in a strong sense; it is a self-generated governing rule that holds the committed individual to the commitment, whether it is a game or a research program to be carried out properly, obtaining correct results, etc. What distinguishes empirical research from other human projects is that in science the skills and concepts of the enterprise itself are in question. However, empirical objects can stand as criteria for the correctness of mundane result due to their "normative authority". "The authority is implicit in the structure of the finding. That normative authority of the discovered objects, derived from but not at all the same as the sui generis normative authority of the constitutive commitment, then takes its effect via the responsible responsiveness of the mundane skills." (P. 343.)

The thirteen essays in the book range over a number of interesting problems in the philosophy of mind and in cognitive science. Haugeland provides careful and detailed surveys of the major views about the issues at hand, making the problems readily accessible to readers of different specialties. The book therefore is an excellent overview of a variety of questions concerning understanding and intelligibility. Haugeland’s own position on these issues is also clearly and rigorously argued, providing engaging and challenging contributions to the field. A thorough critical assessment even of a small number of the arguments Haugeland offers in his versatile book is beyond the scope of this review; many of them would require separate essays in order to receive the attention they deserve. I hope the discussion of Haugeland’s views here showed that the book indeed deserves such attention.

Zsolt Bátori
Rutgers University

Copyright © 2002, Humboldt State University