THE INNER MIND AND THE OUTER WORLD:

Guest Editor's Introduction to a Special Issue on Cognitive Science and Artificial Intelligence

William J. Rapaport

Department of Computer Science

and Center for Cognitive Science

State University of New York at Buffalo

It is well known that people from other disciplines have made significant contributions to philosophy and have influenced philosophers. It is also true (though perhaps not often realized, since philosophers are not on the receiving end, so to speak) that philosophers have made significant contributions to other disciplines and have influenced researchers in these other disciplines, sometimes more so than they have influenced philosophy itself. But what is perhaps not as well known as it ought to be is that researchers in other disciplines, writing in those other disciplines' journals and conference proceedings, are doing philosophically sophisticated work, work that we in philosophy ignore at our peril.

Work in cognitive science and artificial intelligence (AI) often overlaps such paradigmatic philosophical specialties as logic, the philosophy of mind, the philosophy of language, and the philosophy of action. This special issue brings to the philosophical readership of _Noûs_ a sampling of research in cognitive science and AI that is philosophically relevant and philosophically sophisticated.

All of the essays deal with issues of common concern to philosophy, cognitive science, and AI: *intentionality*—the relation of mind to objects in the world; *intentions*—the relation of mind to action in the world; and *reasoning*—the internal workings of the mind, reasoning about both the world and our representations of it. The common theme of the essays—not by design, but neither is it by coincidence—is determining what goes on in our heads when speaking and reading, when planning, and when reasoning. For philosophy, these are clearly central domains of inquiry. For cognitive science, these are crucial if we are to come to an understanding of the nature of the human mind. For AI, these are crucial, since one goal of many AI researchers is to build (models of) minds that must interact with the world.

It seems wrong to call the authors either "linguists" or "AI researchers" or even "philosophers" *simpliciter*. They are all *cognitive scientists*, approaching issues of fundamental concern to all linguists, AI researchers, philosophers, psychologists, etc., using (perhaps, but not necessarily) different methodologies and bringing to bear data from their "home" disciplines as well as from others of the cognitive sciences.[1] In many of the articles, it will be clear, researchers schooled in the methods of their "home" disciplines have adopted (or adapted) the methodology of philosophy.

The six essays in this collection fall into two groups, with interesting overlaps among them. The first three focus on language, mind, and the world. Ray Jackendoff's "The Problem of Reality" considers two ways of understanding the relationship of mind and world: (1) The "philosophical" way sees the world as existing externally to our minds, which can have psychological attitudes (such as knowledge, belief, desire, etc.) towards the world, and it sees sentences of our language as capable of being true descriptions of the world. This is reminiscent of what Hilary

Putnam (1981) has characterized as "externalism" or "metaphysical realism" and what George Lakoff (1987) has characterized as "objectivism". (2) The "psychological" way sees the brain as a physical device to which the world appears in certain ways. This bears a close family resemblance to Putnam's "internalism", Lakoff's "experiential realism", and, perhaps, Jerry Fodor's (1980) "methodological solipsism". Jackendoff assumes the existence of "internal mental representations" defined in a structural or syntactic (a "non-intentional", even non-representational) way, and argues that "the way reality can look *to us* is determined and constrained by the nature of our internal mental representations." His argument proceeds by considering cases where there are no mental representations of external-world phenomena, cases where there are mental representations that differ from that which they apparently represent (as in, e.g., optical illusions), and cases where there are mental representations but no corresponding external-world phenomena.

The investigation of internal mental representations is continued in Don Perlis's essay, "Putting One's Foot in One's Head," a study of causal theories of reference and of intentionality, motivated by issues in knowledge representation and computational semantics (how to deal with meaning in a computational natural-language–understanding system): how are our internal, mental ideas or tokenings of words linked to the external things that they refer to? Perlis examines the "internal mechanisms" for external reference, and suggests that even the primordial dubbing event of causal theories is an "internal matter". He proposes a distinction between "internal reference" and "external reference" to explain how our internal ideas can be meaningful independently of the external world. He thus provides some detail on possible mechanisms for Jackendoff's "psychological" way of understanding the relationship between mind and world. External reference, as found in causal theories, concerns truth conditions, viewing meaning as "public" and "largely outside the head." In contrast, internal reference ("internal 'mental' meaning") concerns "how the agent sees the world, not how the world actually is." Perlis takes internal reference as a relation between a token in a cognitive agent's head and a person-dependent "notional object" (e.g., a visual image or, perhaps, a Meinongian object),[2] arguing that external reference cannot be understood without understanding internal reference.

Jan Wiebe's essay, "References in Narrative Text," is also concerned with the mechanisms of reference. Here, the domain is that of narrative text and, in particular, how references in such narratives reflect a speaker's internal beliefs and point of view. The internal perspective is relevant here, too: First, she is concerned with how the reader determines whether an expression reflects a character's internal point of view or the author's description of the (fictional) "external" world, where the reader has no independent access to either. Second, in order "to understand references in a narrative text, the reader must maintain" an internal mental model of the fictional world, including, recursively, internal mental models of the fictional world's character's internal mental models. Again, issues in linguistics and philosophy of language (as well as literary theory) are motivated by, and inform, research in computational linguistics. In particular, Wiebe provides a computational and cognitive theory of narrative and the ways that cognitive agents understand it. The common theme of the relationship between internal mental content and external "reality" appears here, too: How do certain sentences in narrative ("subjective" ones) express a character's beliefs even when those beliefs may be "false in the fictional world"? There is a computational/processing issue: how to identify these subjective sentences (alternatively put: how to get a computer to read a work of fiction). Her essay nicely illustrates one of the central contributions of AI to philosophy: providing the nitty-gritty details of theories (in this case, semantic theories) expressed as algorithms.

The second group of three papers focuses on reasoning, both theoretical and practical. Phil Cohen and Hector Levesque, in their essay, "Teamwork," and Martha Pollack, in her essay, "Overloading Intentions for Efficient Practical Reasoning," are concerned with planning (the branch of AI that is the counterpart to what philosophers call "practical reasoning"). In particular, Cohen and Levesque look at planning when several actors are part of a team that has a common goal and that, therefore, needs a "joint intention" based on a "shared mental state". As with the essays in the first group, there is a concern with internalism. Here, the focus is on intentions as "internal commitments to perform an action while in a certain mental state." Their essay is an interesting exercise in how the internal and external worlds must be able to interact: An individual's beliefs, plans, and goals will affect and be affected by those of another individual external to the first. Since the actors are part of a team, the notion of mutual belief is central (as it is with Wiebe's paper).

Pollack is concerned with planning in dynamic environments—ones that can change before the planning is finished. She suggests a way to achieve two unrelated goals when an action intended to achieve one can also help achieve the other. She shows how her method can be more efficient in terms of reasoning and action than the "decision-theoretic model in which a complete set of alternatives is first generated and then weighed against one another," even though the decision-theoretic model might produce a more "optimal" plan. Her work is part of a research program in philosophy and AI, with insights and results from each providing data for the other. As Pollack notes at the end of her essay, "philosophical theories can and do matter to the AI researcher."

There is an interesting similarity between Pollack's work and that presented in João Martins and Maria Cravo's essay, "How to Change Your Mind." Both offer sorts of "on-line" or "real-time" reasoning strategies. Martins and Cravo are concerned with how to revise your beliefs rather than to attempt in some optimal, but no doubt impractical, fashion to believe nothing but the truth right from the start. Just as in Pollack's theory, the plan that is adopted need not be the optimal one, but merely one that will suffice and may be worth while adopting, so in Martins and Cravo's theory, the belief that is adopted need not be the best one, but one that can be revised (and may need to be revised if the world changes) and, until revision is needed, may be worth while believing. The internal/external theme reappears: A cognitive agent has beliefs in the form of an internal "model of its environment". As with Jackendoff, the model need not be a complete description of the external world. And as with Wiebe, the model can be a description of a fictional world.[3] The agent's set of beliefs may need to change as the external environment is *perceived* to change (note: it need not actually have changed): Sometimes new beliefs will be added; sometimes old beliefs (or beliefs inferred from old ones) will need to be revised because they are inconsistent with the new ones; and some old beliefs that were merely plausible or tentative conclusions inferred by means of "default rules" (rules with exceptions) now must be retracted in the light of new evidence. The methodology offered by Martins and Cravo is an extension of relevance logic modified to deal with both belief revision (sometimes called "truth maintenance") and nonmonotonic reasoning.

Although not all philosophers may be interested in all of these essays, I think there is something here for everyone. And that was partly my goal in editing this special issue: to show that there is philosophy being done beyond the academic borders of philosophy. To paraphrase Pollack: AI theories, linguistic theories—cognitive science theories in general—can and do matter to philosophy.

NOTES

[1]For a brief overview of cognitive science and an exposition of the idea of applying different methodologies to the investigation of a common problem, see Rapaport 1990.

[2]On Meinongian objects in an AI context, cf. Rapaport 1985 and Shapiro & Rapaport 1987, 1990.

[3]For a related approach to mental models of fictional worlds, see Rapaport 1991.

REFERENCES

1. Fodor, Jerry A. (1980), "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology," *Behavioral and Brain Sciences* 3: 63–109.

2. Lakoff, George (1987), *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind* (Chicago: University of Chicago Press).

3. Putnam, Hilary (1981), *Reason, Truth and History* (Cambridge, Eng.: Cambridge University Press).

4. Rapaport, William J. (1985), "Meinongian Semantics for Propositional Semantic Networks," *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics* (*University of Chicago*) (Morristown, NJ: Association for Computational Linguistics): 43–48.

5. Rapaport, William J. (1990), "Cognitive Science," in A. Ralston & E. D. Reilly (eds.), *Encyclopedia of Computer Science and Engineering*, 3rd edition (New York: Van Nostrand Reinhold, forthcoming); pre-printed as *Technical Report 90-12* (Buffalo: SUNY Buffalo Dept. of Computer Science, May 1990).

6. Rapaport, William J. (1991), "Predication, Fiction, and Artificial Intelligence," _Topoi_ 10: 79–111.

7. Shapiro, Stuart C., & Rapaport, William J. (1987), "SNePS Considered as a Fully Intensional Propositional Semantic Network," in N. Cercone & G. McCalla (eds.), _The Knowledge Frontier: Essays in the Representation of Knowledge_ (New York: Springer-Verlag): 262–315.

8. Shapiro, Stuart C., & Rapaport, William J. (1990), "Models and Minds: Knowledge Representation for Natural-Language Competence," in R. Cummins & J. Pollock (eds.), _Philosophical AI: Computational Approaches to Reasoning_ (Cambridge, MA: MIT Press, forthcoming); pre-printed as _Technical Report 90-10_ (Buffalo: SUNY Buffalo Dept. of Computer Science, May 1990.)