13e CONGRÈS INTERNATIONAL DE CYBERNÉTIQUE

Namur (Belgique), 24-28 août 1992

13th INTERNATIONAL CONGRESS ON CYBERNETICS

Namur (Belgium), 24th-28th August 1992

ACTES

PROCEEDINGS

Namur

ASSOCIATION INTERNATIONALE DE CYBERNÉTIQUE

1992

SYMPOSIUM XV

INVARIANTS OF COGNITIVE SCIENCE : SCOPE, LIMITS, IMPLICATIONS

Président - Chairman : Petros A.M. GELEPITHIS

TRUE INVARIANTS : A RESPONSE TO HERBERT SIMON

By Petros A.M. GELEPITHIS

ABSTRACT

We review some of Herbert Simon's work on invariants of human behaviour. We first define the notions of human invariance, near-invariance and true invariance and then : (i) we briefly argue that two of the proposed laws of qualitative structure, namely, "the principle of bounded rationality", and "the prevalence of heuristic search", do not qualify for the status of human invariants (and in the case of the latter not even of that of near-invariance) ; and (ii) we briefly comment on a couple of quantitative findings which Simon claims to be among the most important cognitive invariants.

INVARIANTS VERITABLES : UNE REPONSE A HERBERT SIMON

RESUME

Nous faisons une critique concise des vues de Herbert Simon sur les invariances du comportement humain. D'abord nous définissons les notions d'invariance humaine, de proche-invariance et d'invariant véritable après quoi : (i) nous argumentons laconiquement que deux des lois de la structure qualitative qui sont proposées par Simon, "le principe de la rationalité bornée" et "la fréquence de la recherche heuristique", ne qualifient pas pour le statut des invariants humains (en particulier, "la fréquence de la recherche heuristique" ne qualifie pas même le statut de la proche-invariance) ; et (ii) nous commentons deux conclusions quantitatives qui, selon Simon, sont parmi les plus importants invariants cognitifs.

1. INTRODUCTION

> "The fundamental goal of science is to find invariants, such as conservation of mass and energy and the speed of light in physics".
>
> (Simon 1990 p. 1)

P.A.M. GELEPITHIS, Kingston University, Kingston upon Thames, KT1 2EE - England.

The motivation for this brief paper draws upon Simon's (1990) article in the Annual Review of Psychology and Simon and Kaplan's (1989) paper in the *Foundations of Cognitive Science* anthology. The major part of our review will concern the 1990 article. Specifically, we discuss two of the proposed laws of qualitative structure (the principle of bounded rationality ; and the prevalence of heuristic search), and we briefly comment on a couple of quantitative findings which Simon claims to be among the most important cognitive invariants.

2. SIMON'S POSITION : A BRIEF CRITIQUE

On the basis that biological systems change (e.g., through learning, natural selection, socially transmitted knowledge etc) Simon and Kaplan (1989 p. 38-39) conclude that biological invariances are limited in time and bounded in space. Accordingly, they suggest to "seek invariants in the inner and outer environment that bound the adaptive processes". (ibid p. 39). This emphasis on adaptivity and the limited nature of, in particular, Psychology's invariants is very clearly expressed in the following quotation :

> "[Psychology's] laws are, and will be, limited in range and generality and will be mainly qualitative. Its invariants are and will be of the kinds that are appropriate to adaptive systems. Its success must be measured not by how closely it resembles physics but by how well it describes and explains human behavior". (Simon 1990 p. 2).

I agree only with the third sentence and the part of the first sentence expressing the view that the laws of psychology will, in the main, be qualitative. The rest of this section is devoted to a critical review of the rest of the above quotation.

First a few definitions and remarks. For a psychological law to be a *human* invariant it must be applicable to all humans and not being limited in time ; for a near-invariant the time requirement is dropped. I would like to stress here that for a near-invariance only the requirement of time applicability is dropped. Our notion of near-invariance is equivalent to Simon's use of the term invariant. His expression "laws of qualitative structure" means a qualitative type of invariant (in Simon's sense of the word). Therefore, psychological 'laws' which may be found applicable to a particular society (and, possibly, only applicable to a particular group within a society) are not considered as near-invariances. I would call such 'laws' group-laws. Finally, a law that holds true for all sentient beings (human, machine, or extraterrestrial) independent of time we shall call an invariant of the sentient space (equivalently (true) invariant). In other words we reserve the term (true) invariant for the type of invariances exemplified in the motto of this paper.

Of course, our definition of human invariants raise important theoretical and practical questions. For example, on what basis could one decide that a particular psychological principle was equally applicable to the citizents of ancient Hellas (Greece) as it is applicable to a modern human being ? On the positive side, establishment of true invariants can provide new insights into problems like the delineation of *homo sapiens*.

We turn now to our review proper. On the basis of the fact that the game of chess is beyond exact computation and heuristic or approximate computational methods have to be introduced for enhancing computer chess playing, Simon

derives what he believes to be one of the most important "laws of qualitative structure"[1]. Namely :

> "Because of the limits on their computing speeds and power, intelligent systems must use approximate methods to handle most tasks. Their rationality is bounded". (Simon 1990, p. 6).

Let us examine closer this statement. Its key element is the notion of intelligent system[2]. If one accepts the PSSH, (and one must be aware as Simon and Newell are, that this is *still* a hypothesis despite being successfully tested over a good range of activities termed 'thinking'), then it follows that the human 'is' a PSS. It does not nevertheless follow that the human brain is *just* (or only) a PSS. That would have been an additional claim and Simon (1990) does not make such a claim. In other words, accepting the PSSH one is forced to accept only that the human brain *can* act as a PSS. So, when Simon says that the rationality of intelligent systems is bounded he should mean that as far as humans are concerned they *can* behave rationally, that is computationally (or formally) within the bounds imposed by the physical limits of computational speed and memory. Subsequently, the quoted statement is true for purely formal systems (an implicit characteristic of any physical symbol system) but is not necessarily true for systems like humans which employ informal as well as formal methods and terms. Therefore, the proposed 'principle' of bounded rationality can be accepted as a *PSS* near-invariant, but can hardly be justified to be proposed as a near-invariant for humans. Is the proposed principle inadequate for PSS invariant status ? This question leads to some interesting issues but to attempt to address it here would take us beyond the scope of this paper.

Simon then goes on and describes "some of the mechanisms used by human bounded rationality" and proposed heuristic search as a basic near-invariance. His argument boils down to the following point : if the task domain has little structure or the structure is unknown to us, we apply domain-independent heuristics (equivalently weak methods). In Simon's words :

> "When intelligence explores unfamiliar domains, it falls back on "weak methods", which are independent of domain knowledge. People satisfice - look for good - enough solutions - instead of hopelessly searching for the best". (Simon 1990, p. 17).

And he furthermore notices that heuristics like means-ends analysis :

> "have been observed as central features of behavior in a wide range of problem-solving behaviors where recognition capabilities or systematic algorithms were not available for reaching solutions without search. The prevalance of heuristic search is a basic law of qualitative structure for human problem solving". (ibid p. 10).

But problem solving behaviour is only one type of human behaviour, behaviour is only one aspect of human life, and even within problem-solving behaviour heuristics is only one of several machanism (Simon actually mentions three).

[1] This is a science-bound notion and refers to those general statements within a science, S, which characterize the essential nature of the S-systems under investigation (Newell and Simon 1976 p. 115).

[2] The reader may consult Newell and Simon (1976) for the original definition of the Physical Symbol System Hypothesis (PSSH), the notion of a Physical Symbol System (PSS), and the associated notion of an intelligent system. Newell (1990) provides an updated and more detailed treatment of these and related issues and Simon (1990) gives a summary of the PSS notion. In this paper we assume familiarity with Newell's and Simon's work on the foundations of Artificial Intelligence and Cognitive Science.

Furthermore, even if one accepts that within *analytical* human problem-solving, - which is actually where human search is employed (ibid p. 9) - heuristic search is the prevalent technique used, this hardly makes it a candidate for a near-invariance. First, because the space of its applicability (the space of analytical, human, problem-solving) is too much constrained. Second, and more important, because other methods like conceptual analysis can be much more significant in terms of the results they may produce. A case in point is Einstein's conceptual analysis of the electrodynamics of moving bodies (actually of the relative motion of a magnet and a conductor) that led to the special theory of relativity (Einstein 1905"1952).

One may be tempted to question the employment of apparently additional criteria like significance in determining the status of a law. The point is that for true invariants such additional criteria are reduntant since true invariants are significant by virtue of their applicability range. It is only when one severely restricts the invariances-space that issues like significance of results obtained need to be incorporated. Simon is silent on this point too. What is left then for heuristic search is nothing more that the usual range of any successful technique ; clearly not enough for it to be elevated to the grade of near-invariances.

I would like to end this review of laws of qualitative structure with a point from the semantics perspective. Heuristics are rules of thumb, that is, rules which experience has shown to be useful. To claim, therefore, that *heuristics* (attention : we did not say a particular heuristic rule or even a kind of heuristic rule) is a near-invariance is like claiming that experience itself is an invariance. It is true of course but it is trivial.

We come now to the quantitative findings proposed by Simon as invariants. They are four (the first three are from Simon (1990), the fourth from Simon and Kaplan (1989) :

1. An act of recognition takes approximately 1 second.
2. The simplest human reactions take approximately (10^{-2} - 10^{-1}) seconds.
3. Human STM can normally hold about a half dozen chunks.
4. Ubiquitous presence of large knowledge bases of the order of 10^5 productions for large expert systems, $5\text{-}10 \times 10^5$ productions for human domain experts.

First of all I accept that the first two findings constitute human invariants. There is nevertheless no ground at all to accept them as true invariants too. It is true that speed and memory limitations may well influence aspects of some of the human cognitive processes. But from this point to the point of claiming that "[t]hey are among the most important invariants that cognitive psychology has discovered,..." (Simon 1990, p. 7) is a claim that needs justification and that Simon does not provide. Similarly, I would not dispute that they *can* account "for many phenomena observed in thinking or learning". (ibid), but accounting for many phenomena is not an adequate reason for a law to be considered an invariance, especially "since *many* other species of thought remain undescribed, a *vast* work of taxonomy and empirical exploration lies ahead". (Simon 1990 p. 4, my emphases). I would, for example, consider limited life span to entail bounded rationality in a more fundamental way that the speed and memory limits quoted above do. Moreover, given that the range of phenomena explained in terms of these findings is quite limited, issues like how fundamental such findings are in entailing the proposed principle of bounded rationality should be taken into account.

Two brief points on the last two of the above quantitative findings. First, Miller's magic number 7 plus or minus 2 is only applicable to humans (not to machines, as a sentient kind of invariance requires, and as one would expect on the basis that Simon's "account will cover the topic of human and computer psychology" (ibid p. 3)). Furthermore, there are important exceptions even within the human space[3]. Hence, human STM capacity may only be hold as a near-invariance.

Second, given that human experts is a small minority among the human beings the ubiquitous presence of extensive domain knowledge only points to its limited significance for humans as a species (at least for the time being and for the foreseeable future). It is, therefore, a group-law rather than an invariance or even a near-invariance.

Summarising, I would say that recognising the difficulties in discovering invariances in Psychology does not justifiy us in glorifying qualitative or quantitative methods or constraints to the status of invariances in the human, machine, or sentient beings space. The claim that :

> "A science of intelligent systems has to be a science of adaptive systems with all this entails for the difficulty of finding genuine invariants". (Simon & Kaplan 1989 p. 43).

is unjustified. Furthermore, it imposes an unnecessary and inaccurate restriction in the range and type of problems addressed by humans. Humans are indeed adaptive organisms *but* they are *adapting* entities too[4]. It is not any more the case that problems are solely created by the human environment. Problems that humans themselves *create* as a result of their thinking, inquiring, and curiosity-driven nature are of equal importance in both shaping the internal human structures and processes and changing their external environment. A human is a physical symbol system but this fact is trivial compared with some of the other characteristics of a human. For example, ability to think in informal terms and therefore not being, eventually, subject to Godel's incompleteness theorems which only apply to humans conceived as purely physical symbol systems.

REFERENCES

EINSTEIN A., (1905*1952). On the Electrodynamics of Moving Bodies. In The Principle of Relativity : A Collection of Original Memoirs on the Special and General Theory of Relativity. Translated by W. Perret and G.B. Jeffery. Dover Publications, Inc.

NEWELL A., (1990), "Unified Theories of Cognition". Harvard University Press.

NEWELL A. and SIMON H.A., (1976), Computer Science as Empirical Inquiry : Symbols and Search. Communications of the ACM, Vol. 13, pp. 113-126.

SIMON H.A., (1990), Invariants of Human Behavior. Annu. Rev. Psychol. Vol. 41, pp. 1-19.

SIMON H.A. and KAPLAN C.A., (1989), Foundations of Cognitive Science. In Foundations of Cognitive Science M.I. Posner ed., MIT Press.

[3] See for example Luria's *The mind of a mnemonist*, performance of patients suffering from impaired STM, or the memory restrictions which frequent accompany old age.
[4] As a matter of fact the adapting aspect of human nature is exponentially increasing in the last century or so, as it can easily be witnessed by the effect of the human actions on our environment.

THE 'SHARED KNOWLEDGE' PARADOX AND THE OBJECTIVITY OF KNOWLEDGE AND BELIEF

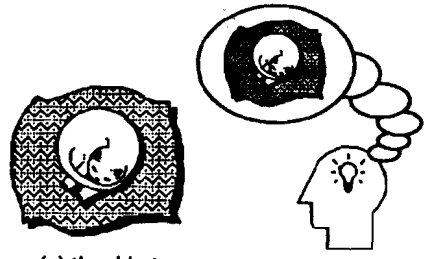
By C. HUTCHISON

1. INTRODUCTION

The rules underlying cultural (including linguistic) behaviour are taken to be cognitive invariants (Chomsky, 1986 ; Spradley & McCurdy, 1972, p. 8ff). Accounts of why this is so (e.g., Chomsky, 1986 ; Jackendoff, 1983) make appeal to the common genetic inheritance and environmental exposure of individuals. I do not find this response satisfactorily resolves the following paradox : if the ultimate locus of the learning process is in individuals rather than in groups, then in what sense can a language or a culture (or indeed any body of knowledge) be said to be *shared* by members of a community ? or again, if my language is really just my idiolect, construed mentalistically (and nothing other than idiolects in that sense are real 'things'), in what sense of the word 'same' do I and my fellows speak the 'same' language ? It may be that the questions are misguided, taking the words 'shared' and 'same' more literally than is warranted by the facts of the matter ; possibly the words should be used more loosely, in which case the 'Shared Knowledge' paradox, as I shall call it, evaporates. I do not believe this to be an acceptable solution (though there is not space in this paper to say why ; see Hutchison, 1992), and below I outline what I think to be novel arguments in support of an alternative account of cognitive invariants in language and culture.

2. KNOWLEDGE AND KNOWING

An important distinction must be made, on the one hand, between that which is known - let's call it the *object* of knowing (see Figure 1a) - and its representation (Figure 1b), and on the other hand, between the representation and the (conceptual) content of that representation (Figure 1c). This is not self-evidently true. It is a common assumption, for example, that knowledge of a language, whatever the form of its physical embodiment in the brain, actually *is* the language, or that the mental representation of the meaning of a word actually is the meaning of the word, that there are no mysterious 'things out there' that they could otherwise be. By analogy, the social and cultural knowledge I draw on in reasoning about the actions and intentions of others is wholly and solely whatever I have learned as an



(a) the object (b) mental/brain state

"The cat sat on the mat"
(c) propositional content of mental state

Figure 1: Knowledge, its object, and its content