Different Types of Intelligence: Bridging the Gap Between Systems Design and Cognitive Theory

Timothy Harrison
University of Bath
Bath, BA2 7AY, UK
Tsh20@bath.ac.uk

ABSTRACT
In this paper, the nature of intelligent systems is considered, and the limitations of current autonomy demonstrated. The potential problems associated with anthropomorphizing these systems are explored, and a potential solution is suggested, in the form of a bridging mechanism that will better relate cognitive theory about knowledge to system interfaces.

Author Keywords
Autonomy, Intelligent Systems, Bridging Mechanisms, Cognitive Modeling

INTELLIGENT SYSTEMS

An artificial intelligence that can converse directly with humans on an equal footing has been a staple of science fiction for many years. From characters speaking directly to a ship's computer on 'Star Trek' as if it were a person, to the blurred human-robot distinctions, complex ethics and questions of sentience in Isaac Asmiov's Robot [2] series, naturalistic human-esq interaction has been the implicit ideal to which technology is expected to aspire.

Intelligent interfaces can be seen as a step towards this goal. However, by its very use, describing something as 'intelligent' is a pejorative term, creating expectations and implying capabilities such as the fictional ones detailed above – intentionally or otherwise. It is important to understand what is meant by 'intelligence' in this context as compared to its human equivalent, and what this means if we are to properly assess and support interaction with these systems. It is the intent of this paper to examine these issues and a theoretically informed manner of developing of these systems is proposed, with specific reference to knowledge management and application as an example.

CURRENT SYSTEMS

The systems that exist at the moment are not what would be thought of, colloquially, as 'intelligent'. Intelligence suggests an ability to make decisions independently and react to dynamic circumstances appropriately – it suggests a degree of autonomy that modern systems do not possess [5]. That is not to suggest that such systems are not cleverly designed, as they frequently are, or that they cannot enhance a user experience and/or efficiency. However it important to note that no system (or interface) can as yet can be declared truly intelligent. Rather, intelligent systems are possessed of a level of behavioural autonomy.

Behavioural autonomy is characterized by the ability to perform and carry out tasks independently, but only under predefined conditions and circumstances, to specified starting and ending positions [5]. These can be rigorously set, or exist as fuzzier operating parameters that allow for greater flexibility, but are ultimately constrained behavior that lacks any true independent thinking or reasoning. Truly intelligent behavior would represent constitutive autonomy – broadly the kind of independent thinking and reasoning seen in humans.

The significance of this distinction lies in understanding what is and is not currently possible for intelligent systems to achieve, and what that means for interacting with them. Interfacing with people is a natural and everyday occurrence; interfacing with artificial autonomy is a different prospect that creates a unique set of challenges. To create a means of interacting with these systems in a meaningful way (an intelligent interface) the interaction itself needs to be better understood and documented.

The consequence of these differences in the way that systems and humans 'think' will be a significant factor in their use and development. These issues can be broadly thought of as capability differences and unrealistic expectations.

Capability differences refer to the intrinsic differences between how machines process information, and how humans do. Artificial systems work on rule-based, digital, parallel processing rules. Humans make decisions through heuristics, in analogue and in serial [7]. These represent fundamental differences in processing that reflect both the nature of information capable of being processed, and the types of decisions that can be made as a result. Specifically, humans are capable of making inferences and appreciating context in a manner that artificial systems cannot. Similarly, however, artificial systems can search through huge reams of data in a manner and with a speed and accuracy humans are thus far simply incapable of matching. Understanding these different capabilities and tailoring the interaction appropriately is an essential task, both for a better interaction and also due to the second point; unrealistic expectations.

Unrealistic expectations refer to potential gaps created between the user's expectations and understanding of the system and its actual capabilities. This isn't simply a theoretical phenomenon, but one that we can see already in modern-day computing use. People will commonly assume that a computer is doing things in a correct manner or in the most efficient way regardless of whether that is the case or not. For example, there are any number of stories of people following sat-nav systems in blind faith, and being confronted with a bridge their truck cannot pass under, or a ferry connection that doesn't run on that day.

These are not simple user errors, where proper use of the technology would have prevented the mishap. Neither are they solely limitations in the software: yes, technology can be developed to address a specific problem, but the point is a broader one. Users invariably develop a certain amount of trust in a system, particularly one that performs the tasks asked of it in an efficient and reliable manner – in itself not a bad thing. Many systems do indeed perform the tasks asked of them in an effective and reliable manner. And yet these systems have limitations and inevitably cannot address all possible scenarios with the adaptive reasoning a human being could. Consequently, they make a decision – take that road – that a human with the same information would realize is a mistake.

The problem is one of increasing anthropomorphism. As systems take on the illusion of true intelligence more, and employ increasingly sophisticated methods of parsing decision making, this property will not simply disappear. Rather it is probable that it will continue to be as likely – if not more so as systems take on the appearance of intelligence – but occurring in increasingly important situations as more and more complex and sophisticated tasks are entrusted to autonomous systems.

Of course, an autonomous system would not be autonomous if it did not have at least the potential to reach an independent conclusion that contradicts the position of the user. Ideally such conflict-based interaction should lead to synthesis rather than conflict. After all, autonomous systems will be using different forms of information in a different manner to humans and vice versa. We should seek to exploit the diversity of information rather than curtail it.

MODELING INTELLIGENCE

Accepting that it is necessary to study and better understand the interaction between user and system, the question that next arises is how to represent this interaction. Turning to psychology cognitive models would seem to present themselves as an intuitive fit: computational representations of though processes. Such initial potential is somewhat dampened, however, by a range of potential solutions and models that have failed to provide intelligent systems: GOMS, ACT-R, SOAR et al [1,4,8]. Do past failures mean this is a theoretical dead end?

The first possibility here is that cognitive modeling offers no decent account of human cognitive function in the first place, and as a result any approach that employs it is following a flawed methodology. It is fair to say that the field does have it share of detractors that would argue this position, but that to take a look at it from a wider perspective, it has offered too many successes to be dismissed so summarily. It has offered quantitatively supported account of aspects of human thought processes and mental structures, as well as re-applicability into many areas such as clinical therapy [4]. To dismiss it as offering nothing on the basis of one failure in this particular area seems largely irrational.

It is the position of this paper that, instead, there are two problems that should be addressed, acknowledgment that it represents an incomplete theory, and development of bridging mechanisms.

It needs to be acknowledged that we are still in the process of mapping and modeling the human mental capacity, to any degree. There has been great progress, and it is not the intention of this paper to downplay the developments made, but it remains true that our understanding is scattershot, incomplete and largely unconnected. To expect a fully-fledged, reliable and comprehensive model to emerge from this is unreasonable, but at the same time the nascent understanding that does exist can still offer much [6]. What is important is to accept the beneficial limits of that pre-emptively, rather than envisioning more than exists.

To this end there must be found a way that incomplete theory can still inform design, and that practical implementations can similarly impart information to theoretical models. Such a problem also exists in the field of clinical psychology, where practical treatments are, by nature, significantly different to the theoretical models that exist – and yet at the same time the theories can be seen to have been beneficial in furthering treatment. A possible solution to this problem has been proposed Barnard [3] in the form of bridging mechanisms.

Bridging mechanisms are constructs that literally bridge the gap between theory and real-world implementations.

In this paradigm, autonomous systems and intelligent interfaces represent raw, theoretical models of knowledge. Even if they do not map directly to theoretical understanding, they still represent practical instantiations based upon theoretical principles. Even if these principles are old, incorrect or based in mathematics and computing rather than psychology.

This is a place where a bridging representation comes in: a theoretical way of coordinating between an imperfect theory and an overly complex real system that defies a comprehensive model.

It could be argued that GOMS et al are, in fact, bridging mechanisms, and in some ways they are in that they are a simplified way of representing cognitive theory for practical use. However in practice they are not. Where GOMS etc al are established, a bridging mechanism should be fluid and dynamic – responsive to new ideas and feedback from both a theoretical end and real-world data updates. It needs to be a guide to applying cognitive theory, rather than a set method in which to achieve that end.

More to the point, perhaps, it is evident from both the failures of cognitive science to offer a distinct step forwards, and the models' barely passing resemblance to theory that something else is required.

BRIDGING THEORY AND PRACTICE

In the case of intelligent systems, this bridging mechanism can take the form of the interface itself, proving a means of interacting that ensures that both sides of the interaction are operating on the same level, and in a manner useful to both. Therefore it remains to identify an area where such interactions can be studied, to provide data with which to begin to construct such a mechanism.

Fortunately such interactions are not uncommon - in fact they occur in the real world every day. It is common to see two people with conflicting bases of knowledge and understanding negotiating to reach a common ground. What is key is to study and understand the means by which this understanding is reached, and how disparate bases of knowledge are coordinated and communicate in order that we can utilize this understanding to improve interaction with virtual systems. Note that this does not mean a direct replication of human-human interaction patterns.

Rather we should seek to gain and understanding of the type of knowledge transferred, the means by which an understanding is reached and information is transferred. Such understanding would inform interaction design, not only from the point of view of how the system responds, but also in how it feeds back understanding, and also how users are prompted to supply information, and of what sort. Currently, systems are commonly built based upon the requirements from the system end; the information that is required for said system to perform the task it is designed for. Just as important, as has been noted earlier, is understanding the requirements and expectations from the human side in order to prevent misconceptions and undue trust in the system.

What is required, ideally, is a knowledge-rich behavior that systems currently perform badly, or not at all. An example might be sorting irregularly named objects that share common properties. As an example, there might exist two files on a computer, parts one and two of a home movie that have been named 'birthdayfunone.avi' and '02bparty.avi'. Human participants would be capable of deriving semantic knowledge from both of these names to conclude that they might be part of a set, despite the fact that they don't share any direct nomenclature conventions. Understanding how this knowledge can be articulated, shared and coordinated (thus forcing it to be explicitly demonstrated for the benefit of a partner), as well as when and where to apply such sorting suggestions is something that modern systems cannot do well – tending instead to sort by name directly, or other file properties rather than semantic meaning.

Providing a model that establishes a framework within which such knowledge-based processing can occur within would represent a significant step forwards. It would also bridge the gap between the expectations of people and the capacity of intelligent systems as they stand.

CONCLUSION

This paper has discussed the nature of intelligence, and shown that careful consideration of what constitutes intelligence is important for appreciating how these systems can be designed. It is clear that there exists the potential for systems to be more grounded in theoretical positions, and that understanding the nature of human interaction stands to help design more efficient ways for interaction between humans and intelligent systems. What has been lacking to date is a way of relating theoretical understanding in a manner that can be made use of in a practical fashion.

REFERENCES

1. Anderson, J.R. , Matessa, M. & Lebiere, C. (1997) ACT-R: A Theory of Higher Level Cognition and Its Relation to Visual Attention *Human-Computer Interaction* Vol 12, pp439-462
2. Asimov, I. (1950) *I, Robot* Gnome Press, New York
3. Banard, P.J. (2004) Bridging between basic theory and clinical practice *Behaviour and Research Therapy,* no.42, pp977-1000
4. Barnard, P.J. (1985) Interacting Cognitive Subsystems: A psycholinguistic approach to short term memory. In A. Ellis (Ed.) *Progress in the Psychology of Language*, (Vol. 2), Chapter 6, London: Lawrence Erlbaum Associates, 197-258

5. Froese, T., Virgo, N. & Izquierdo, E. (2007) Autonomy: A Review and a Reappraisal *ECAL 2007*
6. Jovchelovitch, S. (2007) *Knowledge in Context: Representations, Community and Culture* Routledge, London
7. Lazlo, E & Masulli, I. (eds) (1993) *The Evolution of Cognitive Maps: New Paradigms for the Twenty-First Centuary* Gordon Breach Publishers, London
8. Laird, J.E. , Newell, A., & Rosenbloom, P.S. (1987) SOAR An Architecture for General Intelligence *Artificial Intelligence* Vol 33(1) pp1-64