Center for Advanced Defense Studies
Innovation for Peace

Directorate of Research

Defense Concepts Series
June 2006

Intentional Cognitive Models with Volition

Ammar Qusaibaty and Newton Howard
[ammar.qusaibaty@c4ads.org] [newton.howard@c4ads.org]

Man's intellectual capacity remains an enigma in cognitive studies, as it is both the subject and the means of analysis. If one is to assume quantum-wave dualism in physics, then the state of the world depends on the instruments we use for observation. Just as electrons cannot be simultaneously treated as both particles and waves, the "paradoxical" nature of investigating human cognition bears inherent limitations. Studying cognitive models may, however, be a seemingly less inconsistent endeavor, so long as "contradictions" can be classified.

In this brief exposition, a variety of aspects related to cognitive models are discussed. The authors maintain that modeling the "paradoxical nature" of human cognition remains the greatest challenge. Consciousness aside, models of conscious systems, or rather *conscious* models of conscious systems, are the main objects of exploration. While intentional systems may seem a good starting point for such an exploration, they lack two important constructs: volition and reflexion. Both concepts, and especially volition, unlike rationality for example, are less discussed in cognitive modeling discourse. Although not devoted to volition or reflexion, this work encourages increased research in these areas.

The present investigation is mainly concerned with the ability of cognitive models to formulate constructs that embed in models the ability to self-assign and manipulate goals based on interaction with the environment and especially with other agents. Reflexive models play an important role here. It is proposed that an agent recognizes its individual goals rather than recognizes goals adopted or learned from other actors or goals endowed by the model's designer.

A conceptual "representation" or token for a physical object, human being, organization or process, a model may represent purely physical objects/activities, non-physical objects/activities or a combination of both distributed in time and space. Model design problems may thus emerge by determining the difference between artificial and non-artificial objects or events.

An artificial object such as a vehicle is a closed system or at least contains a closed subsystem. Following Hunt's definition, a system here indicates "a set of mutually dependent variables that take on different values over time." (Hunt 1999: 7) In a closed system, one may accurately predict a state at time *t* based on knowledge of the state at any time prior to *t*. The design of a model for an artificial object/event thus depends on the interdependence of known and chosen variables. A model's limitation is therefore contingent on the number of chosen representative variables, the degree of uncertainty of variable interdependence and the degree of accuracy chosen for such variables. On the other hand, a non-artificial object or activity (phenomena), such as human cognition, is an open system; knowledge of its current state or previous states is insufficient to accurately predict subsequent states. Variables defining such a system are unknown, or known with a great deal of uncertainty. These variables thus bear greater interdependency and consequently more accuracy.

These levels of uncertainty represent classical problems for models in general and specifically for models of non-artificial entities or systems, such as human cognition. At the basic level, a model must describe the state(s), structure and function(s) of what it represents. Structure and function tend to relate to one another in reciprocal fashion: structure enables function and function enables structure. While an extensive review of the philosophical debates concerning these levels of uncertainty is beyond the scope of this exposition, two basic assumptions are here examined.

Any discussion related to models, and especially to cognitive models, assumes first that all activity can be formalized mathematically in the form of predictive rules or laws (epistemological assumption). On the other hand, and as implied by object-recognition theories (Marr 1982; Biedermann 1987) and other connectionist models, reality objectively consists of a set of mutually independent, irreducible or indivisible facts (ontological assumption). The atomistic view implies that reality consists of a pre-defined structure of isolated context-invariant entities. Every situation thus only has one correct response or solution. Influenced by Martin Heidegger and Maurice Merleau-Ponty,[1] Hubert Dreyfus argued however that "a normal person experiences the objects of the world as already interrelated and full of meaning […] There is no justification for the assumption that we first experience isolated facts […] and then give them significance." (Dreyfus 1992: 269-270) Human beings are thus inherently limited in understanding their own behavior in the same way they understand physical objects, especially artificial objects that they have created. Furthermore, there can be no context-free objective prediction of human behavior or cognition.

The argument presented by Dreyfus highlights the methodological divide between representationism and non-representationism as relates to cognitive models and modeling. General representationism views cognition as a form of computation, whereby the mind partially performs tasks by manipulating representations or symbols of objects in the physical world. A cognitive model thus precisely describes computational processes and how they change over time to realize a task. Herbert Simon and Allan Newell's physical symbol system hypothesis provides the necessary and sufficient conditions for a physical system to become intelligent. A symbolic cognitive model can manipulate and compose symbols and symbol structures. A reformulation of the Church-Turning thesis, (Church 1936; Turning 1936) which defines "symbol processing" in a physical symbol system, identifies the key requirement for complex cognition. (Newell 1980; Newell 1990; Pylyshyn 1989)

Fodor and Pylyshyn argue that any valid cognitive theory or model must be able to compose and interpret novel structures based on former structures. The model must therefore exhibit degrees of productivity and systematicity, whereby it produces an unbounded set of novel propositions with finite means. (Fodor and Pylyshyn 1988) Centered on degrees of productivity and systematicity, most debates on cognitive models highlight a model's ability to evolve in processing input to output as input varies within abstract "systematic" structures. A model that can process "John loves Jane," for example, should be able to process an unbounded set of sentences with the format "X loves Y."

Non-representationists argue that productivity cannot be reduced to purely mathematical formalisms even if levels of abstraction and variable constraints are applied to models. Like Dreyfus, they emphasize that, due to its high dependence on interaction with a given environment, human agency is

[1] The work of Heidegger, Merleau-Ponty and others is based on the work of Franz Brentano and Edmund Husserl, who sustained the theory of Intentionality. Brentano in particular was concerned with the relationship of intentionality to consciousness.

primordial. “The situation is organized from the start in terms of human needs and propensities which give the facts meaning, make the facts what they are, so there is never a question of storing and sorting through enormous list of meaningless, isolated ideas,” explains Dreyfus. (1992: 262) According to this view, an agent and its environment are irreducibly coupled. Within their environment, human agents process input according to their goals and context, a process that may be called “situated cognition.”

The situated aspect of cognition has been extensively discussed in cognitive science since the mid-1980s. Applications include “Situated Action,” (Suchman 1987) “Situated Cognition,” (see Clancey 1997) “Situated Artificial Intelligence,” (see Husbands et al. 1993) “Situated Robotics,” (see Hallam & Malcolm 1994) “Situated Activity” (see Hendriks-Jansen 1996) and “Situated Translation.” (Risku 2000) Each application elaborates on the goals and context of input processing. A number of cognitive scientists and artificial intelligence researchers consider situationality the *sine qua non* condition for any form of natural or artificial intelligence. (Ziemke 2001: 163) French philosopher Maurice Merleau-Ponty argues that conscious life—cognition, perception or desire—is subtended by an “intentional arc” that projects an individual’s past, future, human setting and the physical, ideological and moral circumstances of that individual. This intentional arc brings about the unity of sense, intelligence, sensibility and mobility. (Merleau-Ponty 2002: 157) Intentionality may thus be conceived as central to describing human cognition and intelligence.[2]

While views on opposing ends of the cognitive modeling spectrum highlight different requirements, they define the breadth of such a spectrum. In an attempt to formulate necessary requirements for cognitive models, Alan Newell proposed a list of 13 criteria. (Newell 1980, 1990) These functional constraints, also known as “Newell’s test,” (Taatgen 2003) were further distilled into 12 criteria. (Anderson and Lebiere 2003) Inspired by the Turing Test, (Turing 1950) Newell’s criteria elaborate on issues that a cognitive architecture must address in order to provide a plausible basis for an intelligent system.

According to Anderson and Lebiere (2003), this list may be summarized as such:
1. Flexible behavior: behave as an (almost) arbitrary function of the environment;
2. Real-time performance: operate in real time, respond as fast as humans;
3. Adaptive behavior: exhibit rational and effective adaptive behavior;
4. Vast knowledge base: use vast amounts of knowledge about the environment to affect performance;
5. Dynamic behavior: behave robustly when faced with error, the unexpected or the unknown;
6. Knowledge integration: integrate diverse knowledge and make links;
7. Natural language: use natural language;
8. Consciousness: exhibit self-awareness and produce functional accounts of phenomena that re-reflect awareness;
9. Learning: learn from the environment;
10. Development: acquire capabilities through development;
11. Evolution: arise through evolutionary and comparative considerations;
12. Brain realization: be realized within the “brain” (the physical embodiment of cognition).

[2] Throughout this article, intentionality embodies the quality of having intentions and should not be confused with a quality of actions or other distinctions mentioned in the literature.

To achieve human intellectual capacity, a model must implement the first nine criteria. The last three reflect constraints on achieving these functions. (Anderson and Lebiere 2003) Newell's requirements, which reflect representationist views, were subjected to extensive debate and criticism. (Tadepalli 2003; ter Meulen 2003; Wang et al. 2003; Yand & Bringsjord 2003; Yound 2003) A careful analysis of each criterion in terms of design challenges remains critical in cognitive studies. While a detailed discussion of each criterion is beyond the scope of this work, it remains unclear whether intentionality and situationality, as described by non-representationism, are indeed subcategories in Newell's list.

The concept of rationality is not included in one but in many categories from Newell's list. If category 3, for example, assumes that the agent is rational in the sense that a human agent attempts to best satisfy his or her goals, then intentionality may be assumed since "intentionality… is the mother of rationality." (Dennett 1971: 103) As Dennett (1987) explains, individuals have "intentional states" and exhibit appropriate patterns of consistently rational behavior for "intentionality is the character of one thing being 'of' or 'about' something else, for instance by representing it, describing it, referring to it, aiming at it and so on." (Haugeland 1997) Dennett's intentional stance is not only a construct that facilitates predicting behavior; it offers a conceptualization of systems whose behavior rely on beliefs and desires (hopes, fear, etc.), thus augmenting the physical stance[3] and the design stance.[4]

Especially from a computational point of view, the concept of rationality may only be concerned with the "optimal" algorithmic capabilities of manipulating symbols, objects and their attributes for a purpose regardless of where this purpose resides and of who "manages" it. Perhaps the most important aspect of Brentano's intentionality, Merleau-Ponty's intentional arc, or Dennett's intentional stance is self-awareness of relying on beliefs and desires to observe, analyze and act. The stimuli human agents receive from the environment, and especially from other actors, are multi-layered. Self-maintained or managed "information" (beliefs, desires, hopes, etc.) comprise an essential layer. An agent's stimulus may thus have more than just physical implications that require an understanding of physical rules and procedures; it involves manipulating and managing "information" that relates to individual goals and desires and especially beliefs. This specific information may also relate to other actors' goals, desires and beliefs, as stipulated by the "thinking" agent. In addition to manipulating symbols about physical objects and their attributes in the environment, human cognition as it relates to computation, manages "self-assigned" tokens to such objects and their attributes.

This "executive control" of beliefs and goals is thus necessary for deliberate action.

> "To exert executive control, an agent needs to be able to timely initiate, suspend, cease and co-ordinate the execution of adequate intentions, attentively monitor the proceeding execution of the proximal intentions as well as the environmental changes, detect errors and conflicts, actively maintain and sustain the execution of appropriate intentions, and inhibit intervening or distracting processes. An agent exercises his agency to execute his intention by engaging in the mental activity of volition, which makes it possible for the agent to purposefully, intentionally guide his bodily movements in response to changing circumstances." (Zhu 2004: 251)

Volition, as discussed by Zhu (2004) above, provides for a sound application of this executive control.

[3] Applying knowledge of physical laws or phenomena as a method to predict behavior.
[4] Applying knowledge of design, especially of artificial objects created by the individual at hand, as a method to predict behavior.

The philosophical problems of "free will" are caused by three gaps in practical human reasoning and action, as discussed by Searle (2001: 14–15):

1. "Reasons for the decision are not sufficient to produce the decision,"
2. "Decision is not causally sufficient to produce the action,"
3. "Initiation of the action is not sufficient for action continuation or completion."

According to Zhu, (2004) volition bridges these gapes. Unlike rationality and intentionality, volition is neither explicitly nor implicitly included in Newell's list of functional requirements for cognitive models. In their requirements for a theory of intention, Cohen and Levesque implicitly included the concept of volition: "agents track the success of their intentions and are inclined to try again if their attempts fail." (Cohen & Levesque 1990) For Cohen and Levesque, intention is a choice with commitment. Beyond commitment to goals, however, volition is a self-imposed vehicle for ideas to be actions, for the intangible to be real and for the mental to be physical through rationality and choice. If intentionality is the mother of rationality then volition is the mother of intentionality.

As an immediate consequence and in order to formulate an abstraction of both animate and inanimate systems (e.g. man and machine), one may assign a relative degree corresponding to each "gap" of practical human reasoning and action: a V-degree, an I-degree and an R-degree. Since these degrees are relative, they bear temporal and contextual "order" relations and thus overall structure. The formula below illustrates this construct.

Assume a set of two systems x and y, within a context A, where x and y have three order relations according to volitionality, intentionality and rationality respectively. Thus a Cartesian product P x P x P where $P = \{x, y\}$ can be formulated. For a more concrete illustration, assume that $x \leq y$ (volitionality), $x \leq y$ (intentionality), and $y \leq x$ (rationality). While (x,x,x) is non-comparable to (y,y,y) here, an order structure of the overall "context" (where both x and y are part of such a context) can be described as a lattice (see figure 1). Abuse of the notation—for example, (x,y,x)—is denoted by xyx instead.

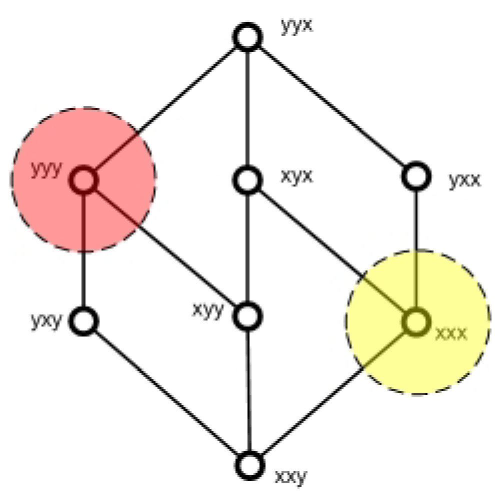

Figure 1 – Volitionality-Intentionality-Rationality Lattice

Such a mathematical structure presents the advantage that every two elements (systems) have upper bounds and lower bounds. Furthermore, every two elements have a least upper bound (supremum) and a greatest lower bound (infimum). The relative degrees of order in volitionality, intentionality and rationality between two systems constitute an order structure beyond these two systems and thus may explain interdependency, cooperation and conflict. Further research in this area is necessary in order to determine the advantage of such mathematical formulations in conveying more subtle relations among systems.

Reflexion[5] is another important implication of volition, in the sense that an agent must assume a self-model and consequently a model of the operating environment, which initiates a fundamental boundary problem between what is and is not included. As a stronger version of Ashby's rule of requisite variety, "any regulator able to confine the fluctuations in the system to be regulated must not only have adequate amounts of variety available to control that system but also be or have a homomorphic representation of that system." (Krippendorff 1986) A regulator may here be understood as a goal-driven system that achieves its goals by regulating the operating environment. Maturana and Varela's theory of Autopoiesis introduces key notions about reflexion, such as: 1) operational closure, 2) component production network (abstraction of metabolism) and 3) spatiotopological unity (individuality and physical borders). (Ruiz-Mirazo & Moreno 2004) The theory's central claim posits that closure or recursivity in an "organization" is generated by the way in which the components and production processes of the system grow intertwined in the context of a complementary relationship between the network and the physical border, (which are simultaneously a mutual condition and result of one another). (Varela et al. 1974) System identity only appears when a coherent set of couplings linking components, processes and flows of energy is established. (Maturana & Varela 1980) This view may suggest that the concept of identity (self) is partially an emergent construct rather than an intrinsic one. Volition partially integrates the emergent self within the capacity of the intrinsic self. Further research may consider the functional relationships between reflexion and volition with an aim to determine whether models of reflexion can be used in part as models of volition. Furthermore, Lefebvre's mathematical formulation of reflexive control (Lefebvre 1977) may prove to be useful tool in this investigation, which requires further research.

Man's intellectual capacity remains an enigma in cognitive studies, as it is both the subject and the means of analysis. If one is to assume quantum-wave dualism in physics, then the state of the world depends on the instruments we use for observation. Just as electrons cannot be simultaneously treated as both particles and waves, the "paradoxical" nature of investigating human cognition bears inherent limitations. Studying cognitive models may, however, be a seemingly less inconsistent endeavor, so long as "contradictions" can be classified.[∞]

In this brief exposition, a variety of aspects related to cognitive models are discussed. The authors maintain that modeling the "paradoxical nature" of human cognition remains the greatest challenge. Consciousness aside, models of conscious systems, or rather *conscious* models of conscious systems, are the main objects of exploration. While intentional systems may seem a good starting point for such an exploration, they lack two important constructs: volition and reflexion. Both concepts, and especially volition, unlike rationality for example, are less discussed in cognitive modeling discourse. Although not devoted to volition or reflexion, this work encourages increased research in these areas.

---

[5] The terminology reflexion is borrowed from Lefebvre's theory of reflexive control (see Lefebvre 1977).

[∞] The authors wish to highlight the reflexive nature of this paper and choose to repeat the introduction.

Bibliography

Anderson, John R. and Christian Lebiere, 2003. "The Newell Test for a theory of cognition." *Behavioral and Brain Sciences* 26.5: 587-601.

Biedermann, I. 1987. "Recognition-by-components: a theory of human image understanding." *Psychological Review* 94: 115-147.

Bringsjord, Selmer and Yingrui Yang. 2003. "Newell's program, like Hilbert's, is dead; let's move on." *Behavioral and Brain Sciences* 26.5: 627.

Church, A. 1936. "An unsolvable problem of elementary number theory." *Amer. J. Math* 58: 345-363.

Clancey, William J. 1997. *Situated Cognition: On Human Knowledge and Compute Representations*. New York, NY: Cambridge University Press.Dreyfus, H. L. 1992. *What Computers Still Can't Do: a critique of artificial reason*. Cambridge, MA: MIT Press.

Cohen, P. R. and Levesque, H. J. 1990. "Intention is choice with commitment." *Artificial Intelligence* 42: 213-261.

Dennett, Daniel C. 1987. *The intentional stance*. Cambridge, MA: MIT Press.

1981. *Brainstorms: Philosophical Essays on Mind and Psychology*. Bradford Books.

1971. "Intentional systems." *Journal of Philosophy* 68: 87-106.

Fodor, J.A. and Z.W. Pylyshyn. 1988. "Connectionism and cognitive architecture, a critical analysis." *Cognition* 28:3-71.

Hallam, John C. and Chris A. Malcolm. 1994. "Behaviour: Perception, Action and Intelligence: The View from Situated Robotics." *Proc. Royal Society Land A* 349:29-42.

Haugeland, John. 1997. "What is mind Design." In Haugeland, John (ed.) *Mind Design II*. Cambridge, MA: MIT Press.

Hendriks-Jansen, Horst. 1996. *Catching Ourselves in the Act – Situated Activity, Interactive Emergence, Evolution and Human Thought*. Cambridge, MA: MIT Press.

Hunt, Earl. 1999. "What is a theory of thought?" In Sternberg, Robert (ed.) *The Nature of Cognition*. Cambridge, MA: MIT Press.

Husbands, Phil, Harvey, Inman and Dave Cliff. 1993. "An Evolutionary Approach to Situated AI." In Sloman, A., Hogg, D., Humphreys, G., Ramsay, A. and D. Partridge (eds.) *Prospects for Artificial Intelligence*. Amsterdam: IOS Press, pp. 61-70.

Krippendorff, Klaus. 1986. *A Dictionary of Cybernetics*. Norfolk VA: The American Society for Cybernetics.

Lefebvre, V. A. 1977. *The Structure of Awareness*. Beverly Hills, CA: Sage Publication, Inc.

Marr, D. 1982. *Vision*. New York: Freeman.

Maturana, H. and Varela, F. J. 1980. *Autopoiesis and cognition: The realization of the living*. Dordrecht, The Netherlands: D. Riedel.

Merleau-Ponty, M. (1945; 1962 trans.) 2002. *Phenomenology of Perception*. London: Routledge.

Newell, A. and H. A. Simon. 1976. "Computer science as empirical inquiry: symbols and search." *Communications of the ACM* 19:113-126.

Newell, A. 1990. *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

1980. "Physical symbol system." *Cognitive Science* 4: 135-183.

Niels A. Taatgen, 2003. "Poppering the Newell Test." *Behavioral and Brain Sciences* 26.5: 621-622.

Pylyshyn, Z. W. 1989. "Computing in cognitive science." In Posner, M. I. (Ed.) *Foundations of Cognitive Science*. Cambridge, MA: MIT Press.

Risku, Hanna. 2000. “Situated Translation und Situated Cognition: ungleiche Schwestern.” In Kadric, M.; Kaindl, K. and Pöchhacker, F. (eds.), Snell-Hornby, Mary (trans.) *Translationswissenschaft. Festschrift für Mary Snell-Hornby zum 60. Geburtstag.* Tübingen: Stauffenburg, pp. 81-91.
Ruiz-Mirazo, K. and A. Moreno. 2004. “Basic Autonomy as a Fundamental Step in the Synthesis of Life.” *Artificial Life* 10.3: 235-259.
Searle, J. 2001. *Rationality in action*. Cambridge, MA: MIT Press.
Simon, H. A. and C. A. Kaplan. 1989. "Foundations of cognitive science.” In Posner, M.I. (ed.) *Foundations of Cognitive Science*. Cambridge, MA: MIT Press.
Suchman, Lucy A. 1987. *Plans and Situated Action: The Problem of Human-Machine Communication*. New York: Cambridge University Press.
Tom Ziemke. March 2001. “The Construction of ‘Reality’ in the Robot: Constructivist Perspectives on Situated Artificial Intelligence and Adaptive Robotics.” *Foundations of Science* 6.1-3: 163-233.
Tadepalli, Prasad. 2003. “Cognitive architectures have limited explanatory power.” *Behavioral and Brain Sciences* 26.5: 622-623.
Ter Meulen, Alice G. B. 2003. “Cognitive modeling of human temporal reasoning.” *Behavioral and Brain Sciences* 26.5: 623-4.
Turing, A. 1950. “Computing machinery and intelligence.” *Mind* 49: 433-60.
1936. “On computable numbers, with an application to the Entscheidungsproblem.” In Davis, M. 1965. *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*. New York: Raven Press.
Varela, F. J., Maturana, H. and R. Uribe. 1974. “Autopoiesis: The organization of living systems, its characterization and a model.” *BioSystems* 5: 187-196.
Wang, Hongbin, Johnson, Todd R. and Jiajie Zhang. 2003. “A multilevel approach to modeling human cognition.” *Behavioral and Brain Sciences* 26.5: 627.
Young, Richard M. 2003. “Cognitive architectures need compliancy, not universality.” *Behavioral and Brain Sciences* 26.5: 628-628.
Zhu, Jing. June 2004. “Understanding Volition.” *Philosophical Psychology* 17.2: 247-273.