“Cognitive Plausibility” in Cognitive Modeling, Artificial Intelligence, and Social Simulation

William G. Kennedy (WKennedy@GMU.Edu)
Krasnow Institute, George Mason University, 4400 University Avenue
Fairfax, VA 22030 USA

Keywords: cognitive plausibility, cognitive modeling, artificial intelligence, social simulation.

Introduction

The claim of “cognitive plausibility” is applied to cognitive models, Artificial Intelligence systems, and social simulations. All three research communities use the term but have different grounds for justifying their use. Can the semantics of the term be rationalized across all three disciplines?

The purpose of this poster is to advance the discussion of the meaning of cognitive plausibility started at the Cognitive Science Society annual meeting in 2008 (W. G. Kennedy, 2008) and continued within the International Journal for Social Robotics (William G. Kennedy, Bugajska, Harrison, & Trafton, 2009).

Different Views of Cognitive Plausibility

The three fields of cognitive modeling, Artificial Intelligence, and social simulation have different views of what constitutes acceptable justification for the use of the desired trait descriptor cognitive plausible.

Cognitive Plausibility in Cognitive Modeling

In cognitive modeling, the focus is primarily on replicating the observed behavior of a single individual and researchers believe theories, experiments, and models matching experimental data are needed to claim cognitive plausibility. With an interest in the make up of cognition, cognitive modeling is focused on experiments that demonstrate overall performance and experiments that isolate components of cognition, such as memory and reasoning. For cognitive modeling, matching human performance data includes matching the errors humans make.

Cognitive Plausibility in Artificial Intelligence

The argument of researchers in the field of Artificial Intelligence is that if the inputs and outputs of the system are comparable to those of humans, then the system is cognitively plausible. The field is less concerned with the cognitive plausibility of the internal components or processes because eventually all the components or processes are implemented in silicon. Hence the black box analogy with no cognitive plausibility claims about the inner working/components/subsystems, i.e., how the outputs are generated. The focus here is on the functional performance of the system. Artificial Intelligence is also not limited to demonstrating the performance of an individual, but is quiet happy to apply multiple and distributed intelligent agents to obtain cognitive performance. Finally, it should also be noted that the goal of AI research is not simply replicating human performance, but understanding the mathematical principles behind it as demonstrated by the building of systems that match and may one day surpass human performance.

Cognitive Plausibility in Social Simulations

The social sciences have the challenge that they cannot conduct experiments on real societies. As a result, social simulations have long relied on functions describing the behavior of rational individuals and behavior of small and large groups as a whole. These formulations go back to difference equations describing the effects of the number of combatants and weapons (e.g., swords and shields or bows and arrows) on one side reducing the number of combatants on the opposing side in each of a series of exchanges (Lanchester, 1916). However, even with the development of much more sophisticated social simulations, the “homo economicus” assumptions of perfectly rational behavior have been criticized by many including Herbert Simon and the community now recognizes a need for better cognitive plausibility in their models of human behavior (Sun, 2006), but is without a definition of what that means.

Common Ground

To find common ground, Nobel prize winner Richard Feynman is instructive. Richard Feynman lectured that “All other aspects and characteristics of science can be understood directly when we understand that observation is the ultimate and final judge of the truth of an idea.” (Feynman, 1998) But cognitive plausibility would then be dependent on “observing” cognition. While we may be getting close to observing cognition directly (Anderson, 2007), simulation has been suggested as a third branch of science, adding to theoretical and experimental branches. Herbert Simon wrote that simulation can be of help to understand the natural laws governing the inner workings of a system from the top down “because the behavior of the system at each level is dependent on only a very approximate, simplified, abstracted characterization of the system at the level next beneath” (Simon, 1969). He also noted that this approach is similar to the foundations for the entire subject of mathematics.

In proposing a unified theory of cognition, Allen Newell proposed several levels within the human cognitive architecture (Newell, 1990) which Ron Sun, and others, simplified to: the sociological level, the psychological level, and the physiological level (Sun, 2006). Finally, John Laird presented an organization to cognitive architectures based on their goal and basis in his plenary presentation at the Cognitive Science Society in 2007. Combining these concepts provides a basis for unifying the various uses of cognitive plausibility for these three areas of research.

Differentiating Cognitive Plausibility

The old problem with the definition of intelligence was that if it was defined in terms of something human did, then no artifact could ever be intelligent and intelligence was not acceptably defined without reference to humans. Similarly, for a cognitive model or system to be worthy of belief, i.e., plausible, is needs to convince us that it is performing cognition. To avoid the arguments about the validity of the Turing Test, a basis for differentiating the uses of cognitive plausibility is proposed here based on observed performance and system levels.

Consider a cognitive system as being made up of one or more layers of systems. I propose defining the cognitive plausibility of any system or layer as:

Proposal for discussion: to be considered "cognitively plausible," a system must be capable of performing as well as humans do on cognitive tasks or be plausibly built on components that have met this test.

To perform "as well as humans do" means matching human performance data. Of course, what it means to match human data is a separate discussion and has been discussed elsewhere, see (Fum, Del Missier, & Stocco, 2007) and (Gluck, Bello, & Busemeyre, 2008) Ron Sun (Sun & Ling, 1997) has proposed three "types of correspondence between models and cognitive [systems]": behavioral outcome modeling (roughly the same behavior), qualitative modeling (same qualitative behavior), and quantitative modeling ("exactly the same" behavior).

Note that this does not address matching human errors in performing cognitive tasks. Being able to match human behavior, both successes and errors, is proposed to be beyond the basic concept of cognitive plausibility. I suggest describing the ability of a system to match human performance including errors as being **"genuinely cognitive plausible"**. Further, to address construction of systems from cognitively plausible subsystems, I propose that cognitively plausibility can be "deep" or "shallow". **"Shallow cognitive plausibility** is cognitive plausibility at only one layer of a cognitive architecture and **"deep cognitive plausibility"** is cognitive plausibility across more than one layer.

For social simulations, cognitive plausibility can be based on using cognitively plausible models for individuals at the next lower level, i.e., for the individuals that make up the society. Using the proposed definition of cognitively plausible, the field of AI can base its use of the term on meeting or exceeding human-level performance. Finally, cognitive model researchers can base their use of the same term on the cognitive plausibility of matching human performance or on a plausible construction of cognitively plausible modules. All fields can clarify their cognitive plausibility as shallow, deep, or genuine. This is the subject of discussion for this poster.

Acknowledgments

This work has been informed by discussions with many individuals including John Anderson, Ken De Jong, Bonnie John, John Laird, Greg Trafton, an anonymous ICCM reviewer, and others over the past year. The work is supported by the Center for Social Complexity of George Mason University and by the Office of Naval Research (ONR) under a Multi-disciplinary University Research Initiative (MURI) grant no. N00014-08-1-0921. The opinions, findings, and conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of the sponsors.

References

Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* Oxford: Oxford University Press.

Feynman, R. P. (1998). *The Meaning of it All*. Reading, MA: Perseus Books.

Fum, D., Del Missier, F., & Stocco, A. (2007). The cognitive modeling of human behavior: Why a model is (sometimes) better than 10,000 words. *Cognitive Systems Research, 8*, 135-142.

Gluck, K., Bello, P., & Busemeyre, J. (2008). Introduction to the Special Issue [on Model Comparison]. *Cognitive Science, 32*(8), 1245-1247.

Kennedy, W. G. (2008). *Cognitive Plausibility of Cognitive Maps and Modeling of Others.* Paper presented at the 30th Annual Conference of the Cognitive Science Society, Washington, DC.

Kennedy, W. G., Bugajska, M., Harrison, A., & Trafton, J. G. (2009). "Like-Me" Simulation as an Effective and Cognitively Plausible Basis for Social Robotics. *International Journal of Social Robotics, 1*(2), 181-194.

Lanchester, F. W. (1916). *Aircraft in Warfare: The Dawn of the Fourth Arm*. New York: D. Appleton & Co.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Simon, H. A. (1969). *The Sciences of the Artificial*. Cambridge, MA: The MIT Press.

Sun, R. (Ed.). (2006). *Cognition and Mulit-Agent Interaction: From Cognitive Modeling to Social Simulation*. Cambridge: Cambridge University Press.

Sun, R., & Ling, C. X. (1997). Computational cognitive modeling, the source of power and other related issues. *AI Magazine, 19*, 113-120.