Kennedy, William G. (2011) The Roots of Trust: Cognition Beyond Rational. In Biologically Inspired Cognitive Architectures 2011: Proceedings of the Second Annual Meeting of the BICA Society, Alexei V. Samsonovich and Kamilla R. Johannsdottir (Eds). IOS Press: Amsterdam. pp 188-193.

The Roots of Trust: Cognition Beyond Rational

William G. KENNEDY[a,1]
[a] *Krasnow Institute for Advance Study*
George Mason University, USA

Abstract. Trust is not simply the result of rational cognition but relies on rational and "beyond rational" cognition. The concept of trust is discussed in terms of its biological, Maslow, and cognitive roots. The cognitive roots rely on the Dual Process Theory, i.e. that there are two types of cognition sometimes called rational and non-rational. Therefore, the roots of trust need a cognitive architecture that implements the Dual Process Theory and involves cognition both rational and beyond rational.

Keywords. Cognitive architectures, rational behavior, Dual Process Theory, System 1 System 2

Introduction

Trust is not simply the result of rational cognition but relies on rational and "beyond rational" cognition. This paper discusses the evolutionary and biologically-inspired roots of trust, its rational and beyond rational components, and the integration of both cognitive components in a cognitive architecture. It is intended to motivate the community to extend cognitive architectures beyond the purely rational cognition to include the other cognitive capabilities necessary to explain trust.

Bertrand Russell quipped, "It has been said that man is a rational animal. All my life I have been searching for evidence that could support this." If man is a rational animal, we certainly do not seem to behave rationally at all time. In fact, we seem to be rational only a small or very small part of the time. As Herbert Simon observed, "Hence, in order to have anything like a complete theory of human rationality, we have to understand what role emotion plays in it." [1] This paper addresses what is beyond the rational side of human behavior and how it is related to the roots of trust.

1. The Concept of Trust

The word **trust** is "officially" defined as a "firm belief in someone or something, acceptance of the truth of a statement without evidence or investigation" [2]. However,

[1] Corresponding Author: W.G. Kennedy, Research Assistant Professor, Krasnow Institute for Advanced Study, George Mason University, 4400 University Drive MS 6B2, Fairfax, VA 22032 USA; E-mail: wkennedy@gmu.edu

the word's definition does not address the reasoning behind the acceptance of the truth of a statement without justification. How is this acceptable? The hypothesis of this paper is that the concept of trust as accepting statements or situations without justification has its roots in biology, Maslow's hierarchy of needs [3], and cognition, both rational and beyond rational.

2. Trust in Biology

Two examples of trust having biological roots are discussed. The first is associated with survival instincts and the second with trust in social circumstances. Survival instincts are behaviors that are performed automatically without being based on learning from experience. They are typically short sequences that are performed based on specific stimulus. An example of these is our irrational reaction to some stimuli, such as the sighting of a spider (the basis of a current smart phone advertisement). The significance of these instinctual behaviors are that they are likely to have survival benefits and to be evolutionarily old, possibly older than our social behavior.

Economists have been studying social behavior, particularly trust, using a game for some time. Recently, neuroscientists have also studied the trust game and have localized brain mechanisms associated with different types of trust. Krueger et al. [4] found two distinct regions were involved in trust. The first was the region associated with representing the mental states of other people, which is related to the formulation of a Theory of Mind for others [5]. Subjects inferred the intentions of others as part of their decision on whether to trust them or not. The other region associated with trust was a limbic region that moderated several aspects of social behavior including reinforcement of goodwill encodings. These results and others demonstrate that trust in a social context has a biological basis.

3. Trust in Maslow's Hierarchy of Needs

Maslow's theory of human motivation [3] proposed a hierarchy of needs as the foundation of his motivation theory. The hierarchy and been discussed extensively but has not been significantly revised. Although "trust" as a word is not found in his classic paper, trust is addressed in the hierarchy at one if not two levels. The first is part of the safety needs, which are just above physiological needs. As he points out, physiological needs must be at least satisfied before safety needs are considered, but safety needs can also be wholly dominating. He also notes that safety can become a dominating immediate goal of an individual and can influence a person's view of the future. Therefore, trust is a major contributor to the perception of safety.

Marlow's hierarchy also includes trust in a person's need for love, affection, and belonging. This need is generally considered to be above both physiological and safety needs, but the levels are not necessarily strictly separate. Trust is included in this level as a major component and not simply in a physical safety sense.

Trust being part of Maslow's hierarchy is important as helping identify the roots of trust as well as the role trust plays in cognition.

4. Trust as Rational and Beyond Rational

In addition to instinctual roots of trust and the role of trust in our hierarchy of needs, trust is addressed cognitively. The cognition associated with trust is discussed in several veins: instinctual evaluations, rational problem solving, evaluation of the results of the problem solving, and with respect to the predictability of the future. Some of this cognition can be considered coldly rational and some instinctual. The importance of this difference will be discussed after the several cognitive veins.

4.1. Gut Feelings

Instinctual evaluations of current conditions determine whether a person is threatened or not. This evaluation has many characteristics. It is fast, nearly instant. It is not consciously done, i.e., it cannot be explained as a series of rational deductions. Its inputs are not obvious, but experiments may identify specific stimuli that are threatening to some individuals (e.g., spiders, snakes, large predators, etc.). These characteristics are more naturally described as gut feelings [6] rather than the result of rational thought [7]. They are also not purely perceptual but involve reactions to perceptions.

4.2. Problem Solving

A second part cognition that is associated with trust is the problem solving associated with the determination of the goal of another agent. We are driven to determine the goal or motivation of another agent. This involves not just watching and following the actions of another agent, but attempting to guess what the purpose of the other agent's actions are. It is a difference between human and other great apes that we determine and mimic the goal-directed behavior, not just the sequential actions of another agent. This problem solving activity is deliberate, time consuming, conscious, and rational.

4.3. Evaluation of Other's Goal

Upon determination of another's goal, we promptly judge whether the goal is threatening or not. Because the problem-solving step takes time, this is a second instinctual evaluation of the situation.

4.4. Predictability and Its Bases

Finally, another evaluation is performed in support of determining the predictability of the other's behavior. There are two methods possible. The first is the integration of the current experience in with previous memories. This process determines if the newly observed behavior consistent with the past behavior or is it significantly different. It also evaluates how this determination effects the predictability of the behavior and its impact on the subject agent. The second method is a judgment or assumption that the other agent is "like-me" and therefore understood [8] including its trustworthiness is known. Besides our inclination to anthropomorphe all agents, we also prefer others who look like us [9].

4.5. Rational and Beyond Rational Aspects of Trust

These several cognitive aspects of trust involve both rational cognition and beyond rational cognition (gut feelings and instinct). Both of the major cognitive architectures ACT-R [7][10] and Soar [11], have models that have addressed emotion [12][13] and However, both are focused on rational cognitive reasoning and, as we noted earlier, rational behavior is only a small part of our behavior. The topic of whether man is a rational or feeling creature is centuries old. A theory gaining credibility is the Dual Process Theory.

5. Rational and Beyond Rational in a Cognitive Architecture

The Dual Process Theory [15][16] postulates both rational cognition and non-rational cognition are both present and necessary to explain behavior. The question of whether people are better described as serial processors or parallel processors is one of the oldest questions in psychology. Several attempts have been made to describe the two processes, but the definitions have varied significantly and typically use terms that carry too much baggage to allow serious discussion. Kahneman [17] in his Nobel Prize acceptance speech suggested using the most neutral terms of System 1 and System 2 to avoid this difficulty.

The System 1 process is generally the evolutionarily old, unconscious, automatic, instinctual, gut feeling, and fast process. System 2 is generally the relatively new, rational, slow, limited process. More details are provided in Table 1. One suggested difference potentially affecting the design of cognitive architectures is that System 1 cognition does not rely on declarative memory but on purely reactive productions instead [12]. These beginnings of clarity on the existence and capabilities of two forms of cognition have begun to be discussed as part of a cognitive architecture [12][18].

Cognitive architectures need to include both the System 1 and System 2 forms of cognition. Soar has included the initial System 1 form of assessment of a situation and used it as the basis for reinforcement learning [11]. ACT-R does not incorporate any System 1 cognition. It might be said that CLARION [18] does in that in includes both implicit and explicit knowledge representation and learning. Learning resulting in "automatic" behavior is similar but not the same as System 1 driven behavior because it is not on the same scale as evolutionary learning [5]. By not including the System 1 cognition discussed above, none of these systems are yet suitable for implementing models of the concept of trust discussed here. This topic was recently discussed at the ACT-R PostGraduate Summer School, July 16-19, 2011, a retreat intended to discuss the future of ACT-R. Presentations are available at: http://cog.cs.drexel.edu/actr2011/program.php.

Table 1. Characteristics of System 1 versus System 2 (from [15]).

General Area	System 1	System 2
Consciousness	Unconscious	Conscious
	Implicit	Explicit
	Automatic	Controlled
	Low effort	High effort
	Rapid	Slow
	High capacity	Low capacity
	Default process	Inhibitory
Evolution	Evolutionarily old	Evolutionarily recent
	Evolutionary rationality	Individual rationality
	Shared with animals	Uniquely human
	Nonverbal	Linked to language
	Modular cognition	Fluid intelligence
Functional Characteristics	Associative	Rule based
	Domain specific	Domain general
	Contextualized	Abstract
	Pragmatic	Logical
	Parallel	Sequential
	Stereotypical	Egalitarian
Individual Differences	Universal	Heritable
	Independent of general intelligence	Linked to general intelligence
	Independent of working memory	Limited by working memory capacity

6. Conclusion: The Roots of Trust

As discussed, the roots of trust involve both System 1 and System 2 cognition. The initial and immediate assessment of a situation seems to be System 1 cognition. The problem solving involved used to determine the goal of another agent appears to be System 2 cognition. The evaluation of the threat of the determined goal of another agent is again System 1. Finally, the evaluation of the history and prediction of the effects of the other agent's actions seems to again be System 2. These steps in evaluating trust therefore involve both System 1 and System 2 cognition.

The purpose of a cognitive architecture was thoughtfully defined by John Anderson as: "A *cognitive architecture* is a specification of the structure of the brain at a level of abstraction that explains how it achieves the function of the mind." [6] page 6. Until cognitive architectures can represent both forms of cognition, trust will be difficult to model and explain.

Acknowledgements

The author appreciates the useful comments by three anonymous reviewers. This work was supported in part by AFOSR/AFRL grant FA9550-10-1-0385 and the Center of Excellence in Neuroergonomics, Technology, and Cognition (CENTEC).

References

[1] H.A. Simon, *Reason in Human Affairs* (1982) Stanford University Press.
[2] *Concise Oxford English Dictionary* (2008) Oxford University Press, Oxford.
[3] A.H. Maslow, A Theory of Human Motivation, *Psychological Review* **50** (1943), 370–396.
[4] F. Krueger, K. McCabe, J. Moll, N. Kriegeskorte, R. Zahn, M. Strenziok, A. Heinecke, and J. Grafman, Neural correlates of trust. *PNAS* **104** (2007) 20084-20089.
[5] R. Dunbar, *The Human Story: A new History of Mankind's Evolution* (2004) Faber and Faber Limited, London.
[6] G. Gigerenzer, *Gut Feelings: The Intelligence of the Unconscious* (2007) Penguin Books, New York.
[7] J.R. Anderson, *How Can the Human Mind Occur in the Physical Universe*, Oxford University Press, Oxford, 2007.
[8] A.N. Meltzof, Imitation and Other Minds: The "Like-Me" Hypothesis. In S. Herley and N. Chater (Eds.) *Perspectives on Imitation: from Neuroscience to Social Science* (Vol 2) (2005), MIT Press, Cambridge.
[9] C. Holden, In Me I Trust, *Science* **297** (2002) 189.
[10] J.R. Anderson, D. Bothell, M. Byrne, S. Douglass, C Lebiere, and Y. Qin, An Integrated Theory of the Mind, *Psychological Review* **111** (2004), 1036-1060.
[11] J.E. Laird, Extending the Soar Cognitive Architecture (2008) In Proceedings of the Conference on Artificial General Intelligence, ISO Press, Memphis.
[12] W.G. Kennedy and M.D. Bugajska, Integrating Fast and Slow Cognitive Processes, in D.D. Salvucci and G. Gunzelmann (Eds) *Proceedings of the International Conference on Cognitive Modeling* (2010), 121-126.
[13] J. Gratch and S Marcella, A Domain-independent Framework for Modeling Emotion. Journal of Cognitive Systems Research **5** (2004) 269-306.
[14] K.R. Scherer, Appraisal considered as a process of multi-level sequential checking. (2001) In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.) *Appraisal processes in Emotion: Theory, Methods, Research.* 92-120. Oxford University Press: Oxford.
[15] J.St.B.T Evans, Dual Processing Accounts of Reasoning, Juegments, and Social Cognition, *Annual Review of Psychology* **58** (2008), 255-278.
[16] S.A. Sloman, The Empirical Case for Two Systems of Reasoning, *Psychological Bulletin* **119** (1996), 3-22.
[17] D. Kahneman, Nobel prize lecture: Maps of Bounded Rationality: a perspective on intuitive judgment and choice, in *Nobel Prizes 2002: Nobel Prizes, Presentations, Biographies & Lectures,* T. Frangsmr (Ed) (2002) 419-499, Almqvist & Wiksell Int., Stockholm.
[18] R. Sun, *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation* (2006), Cambridge University Press, New York.