

A Universal Moral Grammar (UMG) Ontology

Michael DeBellis

Semantics 2018

mdebellissf@gmail.com

<https://tinyurl.com/UMG-Ontology-2018>

What is a UMG?

- First defined by Marc Hauser in his book *Moral Minds*¹
- *“I argue that our moral faculty is equipped with a universal moral grammar, a toolkit for building specific moral systems. Once we have acquired our culture’s specific moral norms... we judge whether actions are permissible,... without conscious reasoning and without explicit access to the underlying principles.”*
- Analogous to Universal Grammar (UG) in Computational Linguistics
- A hypothesized cognitive module that is part of the human genome

UMG Can Be Set Theoretic

- The term Grammar is used only to highlight the analogy with UG
- Indeed, even regarding UG Chomsky has stated: “...the original formulations of transformational grammar were set-theoretic, not graph-theoretic: trees are simply a pedagogical aid”²
- Web Ontology Language (OWL) and Semantic Web Rule Language (SWRL) are excellent tools to model a formal UMG

What is a Cognitive Module?

- First hypothesized by Chomsky (Language Faculty)³, Marr (Vision)⁴ and Fodor⁵
- Chomsky: “[domain specific mental representations] ...what you might think of metaphorically as “mental organs” on the analogy to organs of the body”³
- In the last decade has been widely embraced by the Evolutionary Psychology community as a model to explain many faculties of the human mind: Morality, Theory of Mind, Living Things, Locations and Navigation
- The opposite of the “blank slate” model that hypothesizes one generic mechanism for learning and reasoning

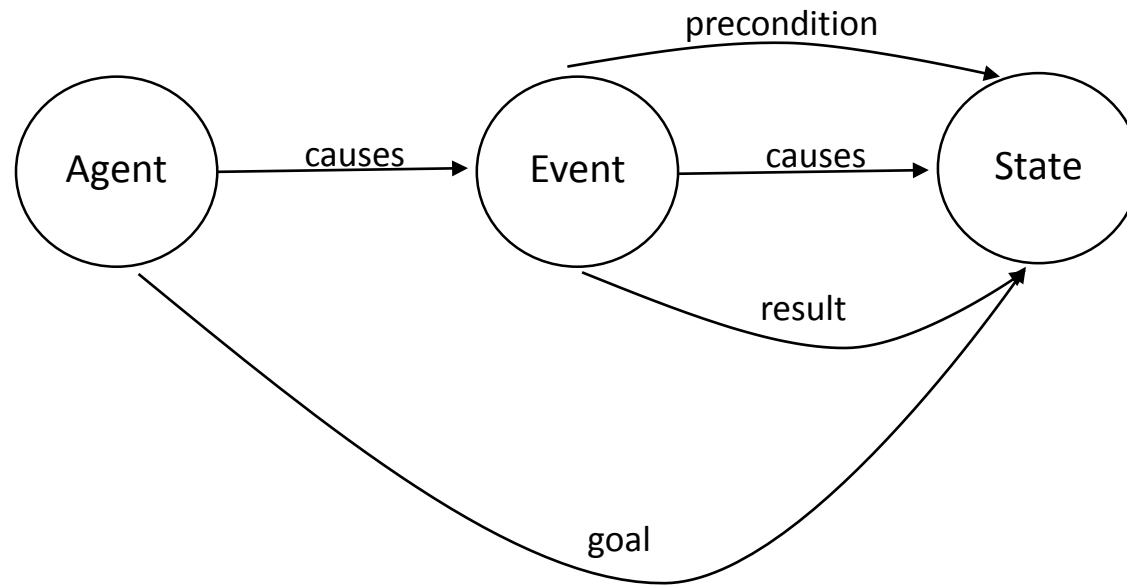
Formal Models of Cognitive Modules

- While there have been formal models of the Language Faculty and Vision there have been no formal models of these other modules except for the work of John Mikhail⁶
- Mikhail has developed a model of a UMG based on English sentences. His model is analogous to “surface structure” in computational linguistics (the syntax of a specific language) where as mine is analogous to deep structure (the hypothesized underlying model used for all natural languages)
- Many researchers describe informal “models” for these other modules
 - Theory of Mind⁷
 - Moral Faculty⁷
 - Living things⁸
- My hope is that this work is a starting point, not just to formally model the Moral Faculty but other modules as well and their interaction
- In order to model a UMG it was necessary to model various aspects of these other modules as well, especially Theory of Mind

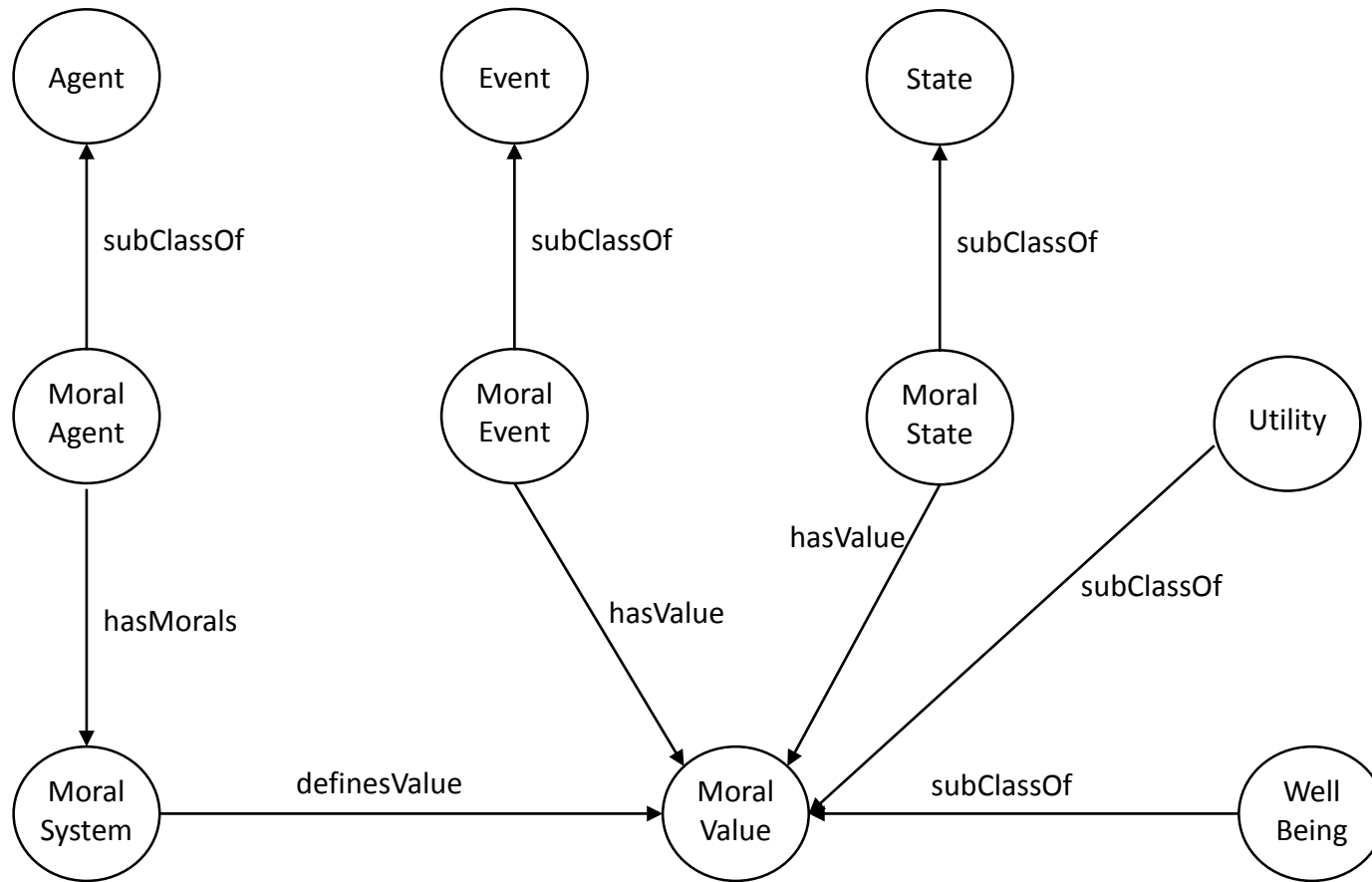
The Starting Point: Theory of Mind

- Describes Agents, Events, Causality
- Evidence for existence in pre-verbal infants⁷
- Used standard AI model for events first developed by Newell & Simon⁹ which has been used for many AI systems such as SHRDLU, SOAR, and the Knowledge-Based Software Assistant (KBSA)

Fundamental Theory of Mind Model



Extending TOM to Create a Moral Model

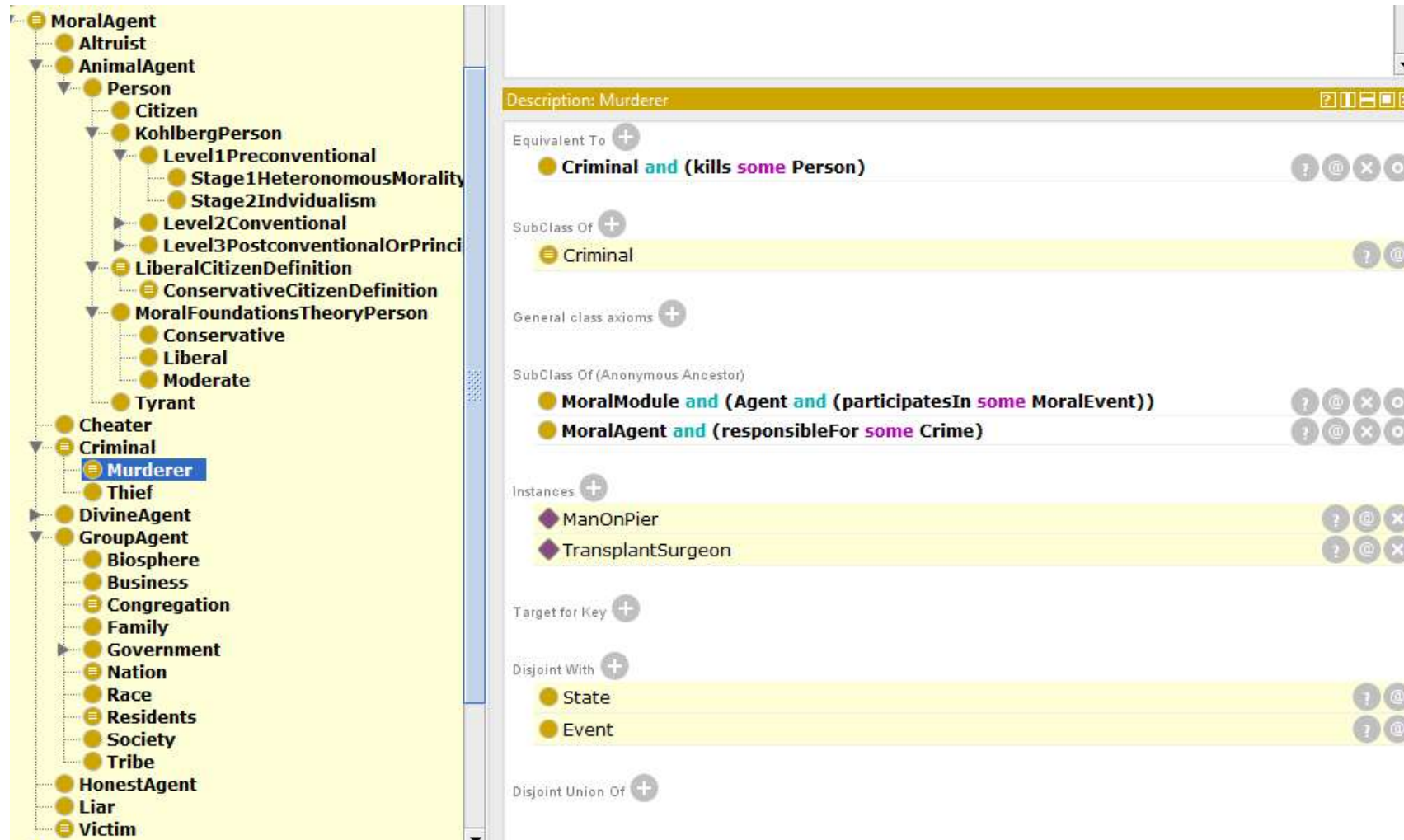


Moral Agent: An Agent that places value on Events and States

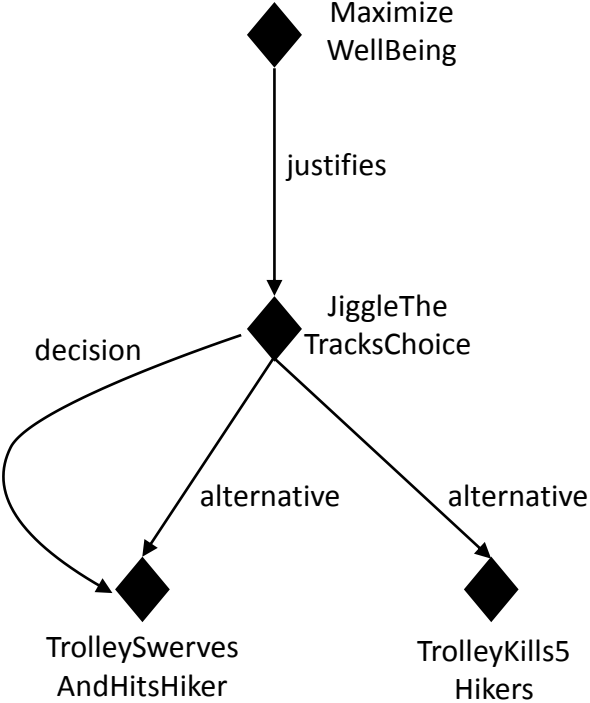
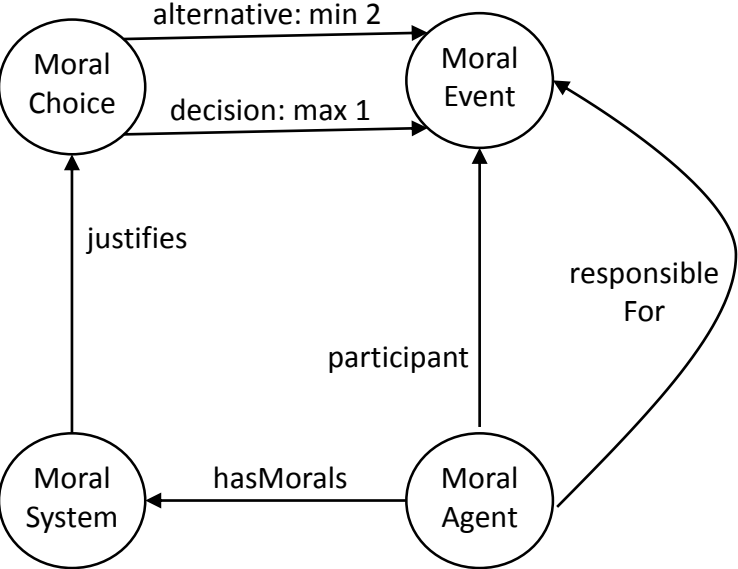
Moral Event: An Event that some Moral Agent places value on

Moral State: A State that some Moral Agent places value on

Example: Moral Agent Subclasses



Important subclass of MoralEvent: MoralChoice



State of Current Ontology

- Implemented over 40 scenarios (Moral Choices) from the philosophical, psychological, anthropological, and biological literature
 - Fischer and Ravizza¹¹
 - Marc Hauser¹
 - Moral Foundations Theory¹²
 - Christopher Boehm¹³
- SWRL Rules that define diverse moral systems
 - Utilitarianism
 - Justice as Fairness
 - Moral Foundations Theory (MFT)
 - Golden Rule
 - Categorical Imperative
 - Fair collaboration/Free Rider identification
 - Religious dogmas
 - Kin selection

Example: Formalization of Utilitarianism

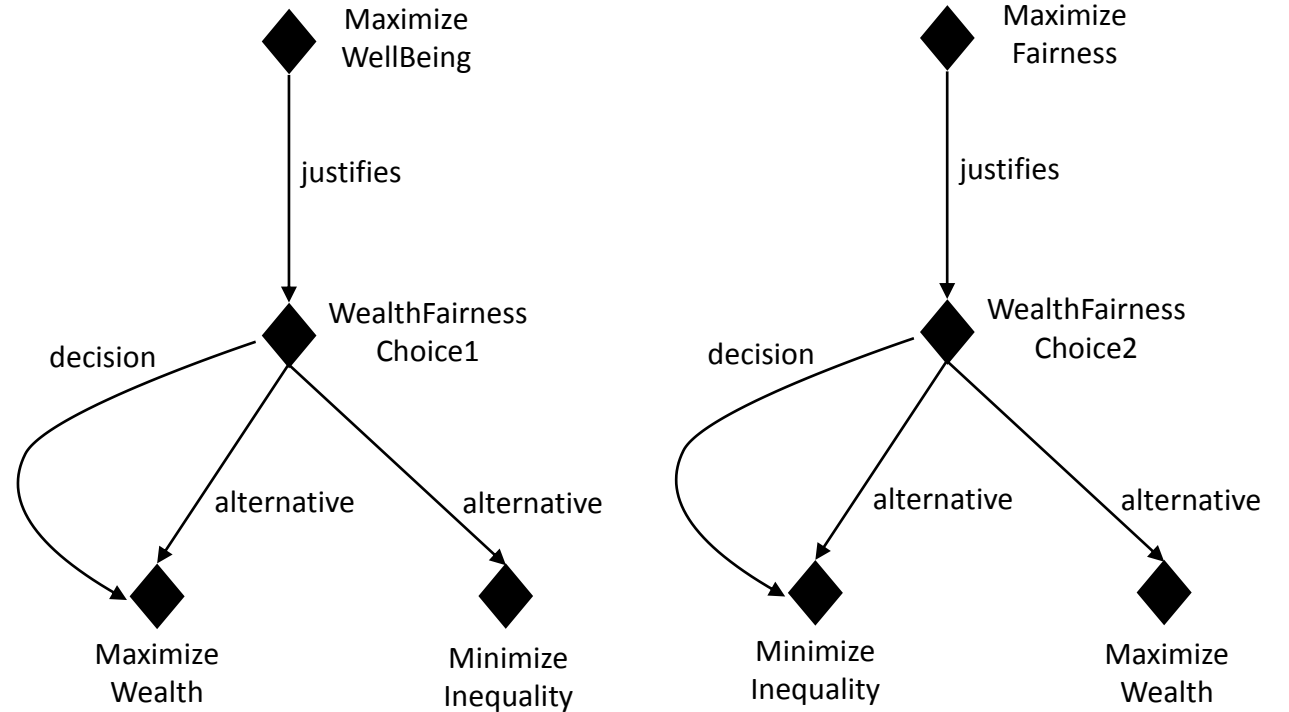
- To begin we need a rigorous definition of what Utilitarianism actually is
- *“A particularly striking expression of the popular misunderstanding... is the famous statement according to which the purpose of social effort is “the greatest possible good for the greatest possible number”. A guiding principle cannot be formulated by the requirement of maximizing two or more functions at once. Such a principle, taken literally is self-contradictory.”* Von Neumann and Morgenstern: Theory of Games and Economic Behavior. p. 11
- I.e., are we
 - A. Maximizing the total amount of well being
 - B. Maximizing the mean well being of the population
- If A then a population of a thousand miserable people is more desirable than a population of a hundred perfectly fulfilled people and one of the most important goals of any utilitarian would be to eliminate birth control and maximize the number of people
- Given the views of Bentham, Mill, and more recent utilitarians such as Sam Harris it seems that maximizing the mean happiness is the correct definition

SWRL Rules for Utilitarianism and Maximizing Fairness

- Utilitarianism: $\text{MoralChoice}(?c) \wedge \text{alternative}(?c, ?a1) \wedge \text{alternative}(?c, ?a2) \wedge \text{result}(?a1, ?r1) \wedge \text{result}(?a2, ?r2) \wedge \text{meanWellBeing}(?r1, ?r1mwb) \wedge \text{meanWellBeing}(?r2, ?r2mwb) \wedge \text{justifiedBy}(?c, \text{MaximizeWellBeing}) \wedge \text{greaterThan}(?r1mwb, ?r2mwb) \rightarrow \text{decision}(?c, ?a1)$
- Maximize Fairness: $\text{MoralChoice}(?c) \wedge \text{alternative}(?c, ?a1) \wedge \text{alternative}(?c, ?a2) \wedge \text{result}(?a1, ?r1) \wedge \text{result}(?a2, ?r2) \wedge \text{standardDeviationOfWellBeing}(?r1, ?r1sdwb) \wedge \text{standardDeviationOfWellBeing}(?r2, ?r2sdwb) \wedge \text{justifiedBy}(?c, \text{MaximizeFairness}) \wedge \text{lessThan}(?r1sdwb, ?r2sdwb) \rightarrow \text{decision}(?c, ?a1)$

Moral Choice: Maximize Fairness or Well Being

Moral Event	Maximize Wealth	Minimize Inequality
Well Being for 95%	-1	4
Well Being for 5%	100	4
Mean Well Being	4.05	4
Standard Deviation of Well Being	22.01	0



The Value of a Formal Semantic Model

- Formal representation of ethical systems (and other problems from philosophy and the “soft” sciences) can add rigor and eliminate the many confusions that result from different researchers using the same term (e.g., altruism) for subtly different concepts
- This initial model is primarily meant to show that formal modeling of this domain is possible and that a wide range of ethical models could be represented by the same model
- However, the current model provides a resolution to one of the most important issues in ethical philosophy: the Is-Ought problem

The Is-Ought Problem

- First described by David Hume: *“In every system of morality... the author... makes observations concerning human affairs; when all of a sudden I am surprised to find, that instead of the usual ... propositions, is, and is not, I meet with no proposition that is not connected with an ought, or an ought not... as this ought, or ought not, expresses some new relation... 'tis necessary that it should be observed and explained; and at the same time that a reason should be given... how this new relation can be a deduction from others, which are entirely different from it.”*
- This is an issue of basic logic. One can't have derivations that result in ought statements without at least one theorem that defines what is being valued.
- Thus the foundation for any rational ethical system must be one or more unproven statements (i.e., axioms) about what has value: fairness, well being, God's will, etc.
- This invalidates a great amount of modern academic moral philosophy (e.g., Foot, Scanlon, Harris) which ignores the Is-Ought problem by essentially appealing to the shared values of the large majority of their community of readers.
- This does not necessarily imply ethical relativism. An analogy is the history of mathematics. Frege and Russell attempted to completely ground mathematics in pure logic but found that it could not be done. This led to ZFC set theory. The ZFC axioms are not arbitrary or based on culture. They were the result of significant effort to determine the minimum axioms that could support conventional mathematics.
- I recently came across a book which although it takes a more informal approach comes to the same conclusion: *Atheist Mind, Humanist Heart* by Lex Bayer and John Figdor. Bayer and Figdor define a small set of axioms that they propose as a foundation for modern secular morality.

Next Steps

- Development of a model based on fair collaboration and free rider suppression based on the work of Christopher Boehm¹³ and others
- Integrate the UMG with Boehm's database of LPA hunter-gatherer tribal norms and sanctions
- Representation of other cognitive modules and integration with a general model of cognition¹⁴ as distributed cognitive modules
- Integration with Jenna in order to develop Java programs that work with the model
 - Represent moral rules for AI agents
 - Develop simulations to test theories regarding altruism and group selection

Bibliography

1. Hauser, Marc (2009). *Moral Minds (P.S.)* HarperCollins. Kindle Edition.
2. Chomsky, Noam (2016). *Why Only Us: Language and Evolution*. With Robert C. Berwick. The MIT Press; Reprint edition (January 15, 2016)
3. Chomsky, Noam (1984). *Modular Approaches to the Study of the Mind*. San Diego State University Press.
4. Marr, David (2010). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. MIT Press.
5. Fodor, Jerry (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. MIT Press.
6. Mikhail, John (2007). *Universal Moral Grammar: Theory, Evidence, and the Future*. *Trends in Cognitive Sciences*. Vol. 11 No. 4..
7. Hirschfeld, Lawrence, A. (1994). *Mapping the Mind: Domain specificity in cognition and culture*. Edited by Lawrence A. Hirschfeld and Susan A. Gelman. Cambridge University Press. Several essays deal with modules of the type discussed in the presentation.
8. Medin, Douglas L. (1999) *Folkbiology*. Edited with Scott Atran. The MIT Press.
9. Newell, Allen (1958). *Report on a General Problem-Solving Program*. With J.C. Shaw and H.A. Simon. *Proceedings of the International Conference on Information Processing*. 30, December 1958. The Rand Corporation. http://bitsavers.informatik.uni-stuttgart.de/pdf/rand/ipl/P-1584_Report_On_A_General_Problem-Solving_Program_Feb59.pdf
10. Miller, George (1960). *Plans and the Structure of Behavior*. With Eugene Galanter and Karl H. Pribram. Holt, Rinehart, and Winston, Inc.
11. Fischer, John Martin (1952). *Ethics: Problems and Principles*. With Mark Ravizza. Harcourt Brace publishers.
12. Haidt, Jonathan (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. Vintage. Kindle Edition.
13. Boehm, Christopher (2012). *Moral Origins: The Evolution of Virtue, Altruism, and Shame*. Basic books.
14. Carruthers, Peter (2006). *The Architecture of the Mind*. Oxford University Press.
15. Yang, Fan (2018). *In Defense of the Commons: Young Children Negatively Evaluate and Sanction Free Riders*. *Psychological Science*. July 16, 2018.