# RESEARCH TRENDS AND OPEN PROBLEMS IN THE FORMALIZATION OF COMPUTER ETHICS

PETROS STEFANEAS

**DEPARTMENT OF MATHEMATICS**

**Shool of Applied Mathematical and Physical Sciences**

**NATIONAL TECHNICAL UNIVERSITY OF ATHENS**

# WHY FORMAL METHODS?

- “Beware of bugs in the above code; I have only proved it correct, not tried it.”

—Donald Knuth

- “Program testing can be a very effective way to show the presence of bugs, but is hopelessly inadequate for showing their absence.”

—Edsger Dijkstra

# FORMAL METHODOLOGIES

- Techniques based on Mathematics and Logic
- Specification, Design, and Verification of Software and Hardware Systems
- Each has its own semantics
- Z, OBJ, VDM, CASL, B-Method, Petri Nets

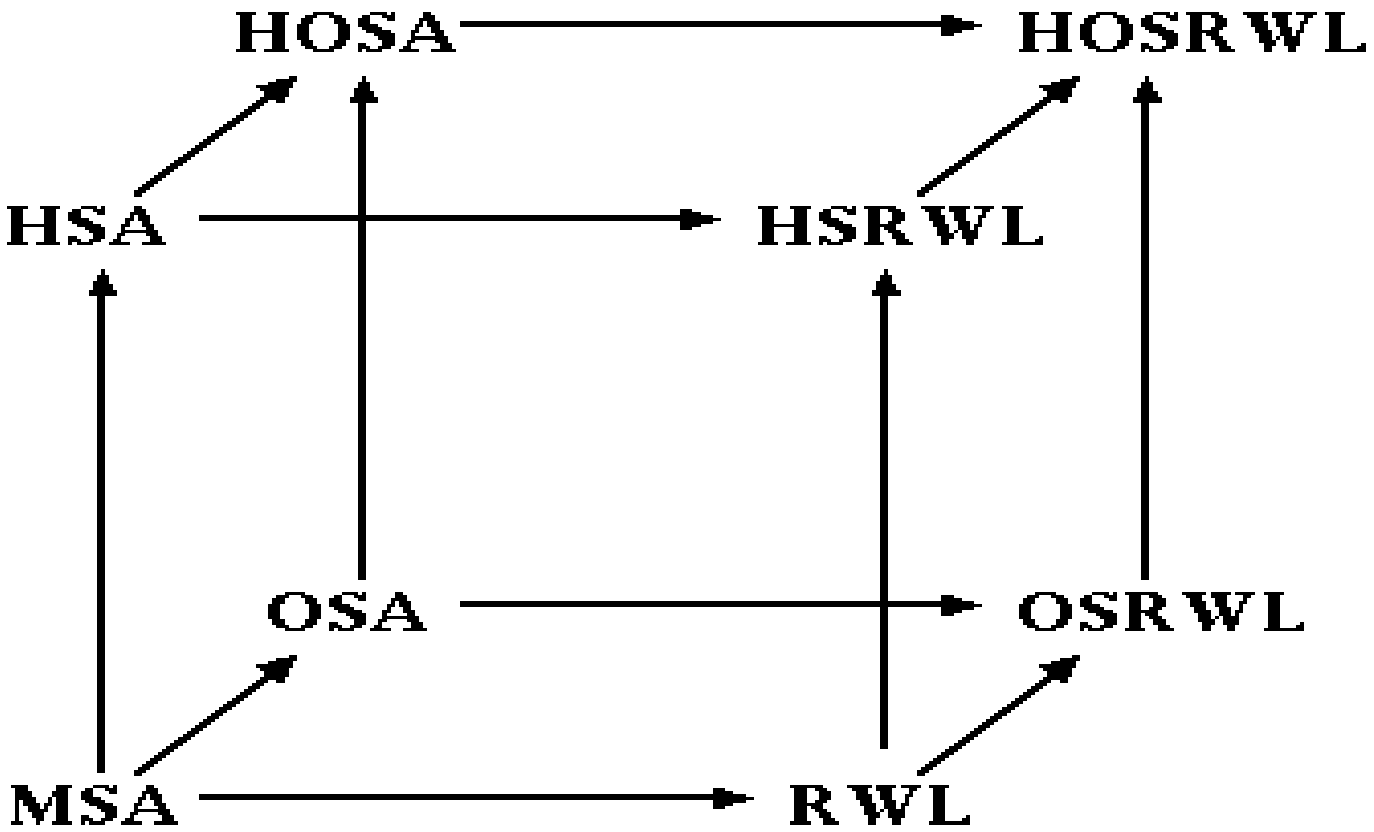
# VERIFICATION

- Input: A specification and a desired ethical property
- Output: “Yes, the property is valid” or “The property is not valid”
- **Very important for computer ethics**

# EXAMPLES OF SPECIFICATION LANGUAGES

- **OBJ Family** (OBJ, Maude, CafeOBJ, FOOPS, 2OBJ, Eqlog)
- ISO: Estelle (Extended Finite State Machine Language) & LOTOS (Language Of Temporal Ordering Specification)
- CCITT: SDL (Specification and Description Language)
- VDM
- Z

# The CafeOBJ Cube: Any Extension?



CS ETHICS SENSITIVE APPLICATIONS

# SOFTWARE – HARDWARE CO-DESIGN

**SOFTWARE SPECIFICATION LANGUAGES**

↑

Higher level languages

↑

Assembly language

↑

Machine language

**HARDWARE SPECIFICATION LANGUAGES**

↑

Hardware Design Languages

↑

Register-transfer language

↑

Gate & transistor level

# Computer Supported Computer Ethics

- DEAL: makes use of recent research in deontic, epistemic and action logic, and on recent research in computer implementations of modal logic
- Athena: mechanized multi-agent deontic logics - vehicle for engineering trustworthy robots.
- Mechanically checked proofs can serve to establish the permissibility (or obligatoriness) of agent actions, and such proofs, when translated into English, can also explain the rationale behind those actions.
- Logic is necessary for valid computer supported computer ethics

# MOTO

WE NEED BOTH

LOGIC & ETHICS

IN COMPUTER SCIENCE:

EVEN FOR PRACTICAL APPLICATIONS

# CS: LOGIC & ETHICS

- CS LOGIC: ***COMPUTABILITY, SEMANTICS OF PROGRAMMING AND SPECIFICATION LANGUAGES, FORMAL METHODOLOGIES, AI APPLICATIONS, AUTOMATED THEOREM PROVING, COMPUTATIONAL LINGUISTICS, SEMANTICS AND VERIFICATION OF SOFTWARE AND HARDWARE SYSTEMS.***
- CS ETHICS: ***PRIVACY, ACCESIBILITY, WORK ETHICS, FAIRNESS, COMPUTER CRIME, SOCIAL ASPECTS.***

# COMPUTER SCIENCE LOGIC

- 1930-50: TURING MACHINES, AUTOMATA, COMPUTABILITY, LAMBDA CALCULUS.
- 1960-80: SEMANTICS OF DECLARATIVE LANGUAGES, ALGEBRAIC SPECIFICATIONS, AUTOMATED DEDUCTION.
- 1990-TODAY: LOGICAL FRAMEWORKS, LOGIC INDEPENDENT COMPUTER APPLICATIONS.

# COMPUTER ETHICS

- B. Russell
- 1940 & 50: Norbert Wiener
- 1960: Donn Parker
- 1970s: Weizenbaum, Maner
- 1973: Code of Professional Conduct for the Association for Computing Machinery
- 1980: Moor
- 1990s: Ethics and the Internet
- 2000s: Privacy

# Cybernetics: or control and communication in the animal and the machine

**It has long been clear to me that the modern ultra-rapid computing machine was in principle an ideal central nervous system to an apparatus for automatic control; and that its input and output need not be in the form of numbers or diagrams. It might very well be, respectively, the readings of artificial sense organs, such as photoelectric cells or thermometers, and the performance of motors or solenoids ... . we are already in a position to construct artificial machines of almost any degree of elaborateness of performance. Long before Nagasaki and the public awareness of the atomic bomb, it had occurred to me that we were here in the presence of another social potentiality of unheard-of importance for good and for evil. (pp. 27-28)**

# PRIVACY: THE KEY ISSUE

- Freedom of Information Act - FOIA (1966)
- Privacy Act (1974)
- Privacy and Anonymity
- 95/46/EK & 97/66/EK
- Greek law N. 2472/1997

# WHY A SEMANTICS FOR CS ETHICS?

- MANY POSSIBLE APPLICATIONS (LANGUAGE DESIGN, SYSTEMS SPECIFICATION, NEW ALGORITHMS)
- COMPUTER SUPPORTED COMPUTER ETHICS
- MORE THAN ONE LOGIC INVOLVED

# CS LOGICS FOR CS ETHICS

- DEONTIC LOGIC
- EPISTEMIC LOGIC
- ACTION LOGIC
- HYBRID SYSTEMS
- ABSTRACT MODEL THEORY: AN INSTITUTION FOR CS ETHICS?

# CS LOGICS FOR CS ETHICS

| Deontic | Action | Epistemic |
|---|---|---|
| The right | To get | Information |
| The obligation | To see to it that | Others know |
| The permission | To let someone | Know |
| Duty | To prevent people from | Believing falsehoods |
| The right | To remain | Ignorant |

# EXAMPLES OF SENTENCES

- **If John has an IP right in a particular piece of information *X*, then Peter ought to have permission from John to acquire, process or disseminate *X*.**
- **If information *X* is about John and if Peter does not have *X* then Peter is not permitted to acquire *X* without John's consent. If he does have *X*, then he is not permitted to process or disseminate it without John's consent.**
- **If *A* is informed about *X*, then all ought to be informed about *X*.**
- **If John has an information responsibility regarding *X*, then John has an obligation to see to it that specific others have access to information *X*.**
- **Agent *A* in informational context *C* sees to it that Agent *B* believes that *p*, or *A* informs *B* that *X***

# THE THEORY OF INSTITUTIONS

- **What is a logic?**
- **Logic independent computer science**
- **Truth is invariant under change of notation.**
- **1990: Goguen & Burstall.**
- **Today: more than 2000 papers and 5 very large scale computer projects**
- **Previous work:**
  - **Algebraic specifications - EQL**
  - **Abstract model theory (Barwise)**
  - **Categorical logic(categorical logics)**
- **Analogies:**
  - **Group theory: $(N,+)$, $S_n$, κλπ.**
  - **Theory of institutions: EQL, FOL, SOL, κλπ.**

# DEFINITION OF INSTITUTIONS

An Institution **I** consists of:

1. A category **Sign** (of signatures),
2. A functor *Sen*: **Sign** $\rightarrow$ **Set** (of sentences),
3. A functor **Mod**: **Sign** $\rightarrow$ $\mathbf{Cat}^{op}$ (of models),
4. A relation $|=_{\Sigma} \subseteq |\mathbf{Mod}(\Sigma)| \times Sen(\Sigma)$, for every $\Sigma \in |\mathbf{Sign}|$ ($\Sigma$-satisfaction), such that for every morphism $\varphi: \Sigma \rightarrow \Sigma'$ of **Sign,** it holds:

$m |= Sen(\varphi)(e)$ if and only if $\mathbf{Mod}(\varphi)(m) |= e$ , for every $m \in |\mathbf{Mod}(\Sigma')|$ and $e \in Sen(\Sigma)$ (satisfaction condition)

# KRIPKE INSTITUTIONS

- POSSIBLE WORLDS SEMANTICS FOR ABSTRACT INSTITUTIONS
- CATEGORIES OF KRIPKE INSTITUTIONS
- **COMBINATIONS – HYBRID SYSTEMS**
- **GROTHENDIECK KRIPKE INSTITUTIONS: A LOGICAL FRAMEWORK FOR CS ETHICS**

# SPECIFICATION LANGUAGES: ALGEBRAIC OPERATORS

- Sum: $T_1 + T_2$
- Transformations for every $\varphi: \Sigma \rightarrow \Sigma'$ and $T = (\Sigma, E) : \varphi * T = (\Sigma, (\varphi(E))^{\bullet})$
- Information hiding: for every $\Sigma'$ and $T = (\Sigma, E) : \Sigma' \bullet T = (\Sigma \cap \Sigma', E \cap Sen(\Sigma'))$
- Expressions

# CURRENT TOPICS OF RESEARCH

- ROBOT ETHICS
- PRIVACY & OTHER LEGAL ISSUES (REGULATION)
- BLOCKCHAIN & SMART CONTRACTS
- ETHICAL THEORY OF INFORMATION
- SEMANTICS OF ETHICS & ARGUMENTATION
- ZERO-KNOWLEDGE PROOFS

# Aris Arageorgis (+2018)