Computational Ethics

Alicia I. Ruvinsky
University of South Carolina, USA

INTRODUCTION

Computational ethics is the integration of computer simulation and ethics theory. More specifically, computational ethics is an agent-based simulation mechanism that takes a computational perspective to ethics theory. This approach uses computer modeling and multiagent systems to generate societies of agents capable of adopting various ethical principles. The principle adopted by an agent will dictate its moral action in response to a moral dilemma. By simulating the agents' application of ethical principles to moral dilemmas and observing the resulting moral landscape of a group of affected agents, we are better able to understand the social consequences of individual ethical actions.

Chung (2004) describes simulation modeling and analysis as "the process of creating and experimenting with a computerized mathematical model of a physical system" (pp. 1-2). A significant advantage to simulation modeling is that once developed, various configurations of the variables comprising the simulation may be explored without incurring the expense or disruption elicited by real-world experimentation (Banks, 1998). It is important to remember that simulations provide a *descriptive* assessment of the system under its current configuration, not a *prescriptive* one. Simulations give an understanding of how the system's configuration relates to the system's behavior, not how to set the configurations so that a certain behavior is elicited (Trick, 1996).

Ethics may be defined as individual principles of conduct or as societal guiding philosophies.[1] The difference between these definitions is a matter of granularity. Ethics as principles of conduct characterize the high granularity of individual agent actions. At a lower granularity, where individual actions are overshadowed by the society's behavior, ethics becomes a societal guiding philosophy. This dichotomy begs the question of how the ethics involved in a principle of conduct manifests itself in a guiding philosophy. Computational ethics attempts to model an ethical system with the intent of observing the dynamics of the system. Hence, in modeling and analyzing a system, computational ethics is exploring the relationship between individual ethical actions and their contributions to the evolution of a large-scale emergent ethic.

Axelrod (1997) gives an example of an emergent ethical system. During the trench warfare of World War I, both sides began to exhibit restraint in killing the enemy, only shooting in retaliation. This group-level behavior was a result of the entities' individual ethics of not wanting to kill the enemy and wanting to defend themselves. How did these individual ethics come about? How does an individual's ethical action affect its neighbors' ethics? How many members of the group must exhibit this ethic before a social ethic emerges? A computational ethics simulation inspired by this episode would assist in exploring these questions.

Computational ethics provides a mechanism for experimenting with and testing social ethical theory. As such a tool, computational ethics could significantly enable a means of facilitating quantitative research in ethics. These simulations may be configured to test proposed theoretical frameworks, allowing for a unique analysis of individual ethical principles and the moral interrelationships that may arise between an agent and its society. Computational ethics provides a mechanism for exploring the consequences of individual moral actions on the emerging social ethic. These consequences in turn may expose hidden ethical dilemmas not foreseen in the original analysis. Discovering the manifestation of secondary (or deeper) dilemmas while a society is attempting to resolve the original dilemma may provide insight into the nature of the particular ethical configuration of the society, namely the kinds of solutions this society generates (Surowiecki, 2004).

Computational ethics addresses the individual behavioral manifestations of an ethic, as well as the emerging social consequences to which individual actions contribute. Computational ethics can be used to explore how multiple individual agents interact with each other as well as the agent society with regard to a moral dilemma, thereby providing a means of

analyzing the evolution of an emergent ethical system of the agent society. As Moss states (2001), "For the social simulator, the issue of how society can impact upon individual behavior is at least as important as how individuals impact on society" (p. 2). As an area of research, computational ethics provides invaluable mechanisms for studying the nature of computational worlds in which certain ethical principles prevail and other worlds in which these same principles may be extinguished. This article provides a background to the conceptualization of computational ethics, followed by a discussion of its implementations and future trends.

BACKGROUND

A natural inclination for human curiosity concerning ethics stems from the many varying and often contradicting moral stances that exist within any group of people, even among close associations such as families. This variance has elicited the exploration and definition of models of morality within the field of ethics theory. The systematizing of morality by defining mathematical models based on formal ethics theory models is useful in a computational sense. These moral systems may be represented computationally and simulated within a multiagent system. This is the backbone of computational ethics.

Ethics

In enabling an agent to adopt an ethic, various simulations may be generated to explore the effects of individual morality on a society. An ethic is a moral framework characterized by rights, liberties, and duties, which are parameters in an ethic model. A *right* may either be a *claim right,* in that it is an agent's opinion about another agent's behavior, or it may be a *liberty right,* which is a right that an agent uses to justify its own behavior. A claim right manifests itself as an external social pressure that an agent senses when making an ethical choice. An agent feels[2] judged by the claim rights of other agents. A liberty right would be used by an agent to justify its action as ethical, making the action immune to social coercion or criticism (Van Wyk, 1990).

Liberty is freedom or autonomy. In a computational agent world, liberty is the means by which an agent interprets the forces in its environment and in its society. If an agent has little liberty, then there exists a controlling entity, such as another agent or the society in social contract ethics. With a large amount of liberty, an agent is more independent[3] in its moral decision making. An ethical theory allows for various liberties, as well as denies other liberties to the agent[4] (Van Wyk, 1990).

A *duty* is either an obligation or responsibility. An *obligation* is when an agent has no moral alternative other than to perform the action represented by the obligation. Not to perform this action would be morally wrong. Obligations can be either directly or indirectly created. A directly created obligation is one in which the agent enters into the obligation at its own volition. For example, an agent has the directly created personal obligation to satisfy its own hunger. An indirectly created obligation results from exterior forces such as social pressure. When these exterior forces become too great, then the agent must behave in the obligated manner.

A *responsibility* is a duty to produce a desired result, or at least contribute to the eventual realization of that result. The agent is not required to perform any specific action, but simply to ensure that its actions contribute to the desired result. For example, Bob has the responsibility to eliminate world hunger. There is no specific action that is required of Bob by this responsibility; however, he reasons that feeding another hungry person will decrease hunger in the world, thereby contributing to the realization of his responsibility. Bob realizes that he "ought" to feed the hungry person, but he is not obligated to do so. If he is occupied in other actions, such as obligations (e.g., satisfying his own hunger), then he may not actually feed the hungry person, though he does acknowledge his responsibility to do so (Van Wyk, 1990).

Moral Dilemmas

With an ethical framework in place, we now consider the moral dilemmas facing the agents. The relationship between individual ethics and social ethics may be explored via moral dilemmas defined in the field of game theory. A prime example of such a moral dilemma is the *prisoner's dilemma* (PD), which can be seen as a moral conflict between duty to oneself and duty to another. The individual actions of rational agents are unable to establish an emergent cooperative