

Introduction

Social and economic interactions today increasingly feature a new category of being: the artificial agent. It buys and sells goods; determines eligibility for legal entitlements like health care benefits; processes applications for visas and credit cards; collects, acquires, and processes financial information; trades on stock markets; and so on. We use language inflected with intentions in describing our interactions with an artificial agent, as when we say, “The shopping cart program wants to know my shipping address.” This being’s competence at settling into our lives, in taking on our tasks, leads us to attribute knowledge and motivations, and to delegate responsibility, to it. Its abilities, often approximating human ones and sometimes going beyond them, make it the object of fear and gratitude: it might spy on us, or it might relieve us of tedium and boredom.

The advances in the technical sophistication and autonomous functioning of these systems represent a logical continuation of our social adoption of technologies of automation. Agent programs represent just one end of a spectrum of technologies that automate human capacities and abilities, extend our cognitive apparatus, and become modeled enhancements of ourselves. More than ever before, it is coherent to speak of computer programs and hardware systems as agents working on our behalf. The spelling checker that corrects this page as it is written is a lexicographic agent that aids in our writing, as much an agent as the automated trading system of a major Wall Street brokerage, and the PR2 robot, a prototype personal robotic assistant (Markoff 2009). While some delegations of our work to such agents are the oft-promised ones of alle-

2 / *A Legal Theory for Autonomous Artificial Agents*

viating tedious labor, others are ethically problematic, as in robots taking on warfare roles (Singer 2009). Yet others enable a richer, wider set of social and economic interconnections in our networked society, especially evident in e-commerce (Papazoglu 2001).

As we increasingly interact with these artificial agents in unsupervised settings, with no human mediators, their seeming autonomy and increasingly sophisticated functionality and behavior raise legal and philosophical questions. For as the number of interactions mediated by artificial agents increase, as they become actors in literal, metaphorical, and legal senses, it is ever more important to understand, and do justice to, the artificial agent's role within our networks of social, political, and economic relations.¹ What is the standing of these entities in our socio-legal framework? What is the legal status of the commercial transactions they enter into? What legal status should artificial agents have? Should they be mere things, tools, and instrumentalities? Do they have any rights, duties, obligations? What are the legal strategies to make room for these future residents of our polity and society? The increasing sophistication, use, and social embedding of computerized agents make the coherent answering of older questions raised by mechanical automation ever more necessary.

Carving out a niche for a new category of legal actor is a task rich with legal and philosophical significance. The history of jurisprudence addressing doctrinal changes in the law suggests that legal theorizing to accommodate artificial agents will inevitably find its pragmatic deliberations colored by philosophical musings over the nature and being of these agents. Conversely, the accommodation, within legal doctrines, of the artificial agent will influence future philosophical theorizing about such agents, for such accommodation will invariably include conceptual and empirical assessments of their capacities and abilities. This interplay between law and philosophy is not new:² philosophical debates on personhood, for instance, cannot proceed without an acknowledgment of the legal person, just as legal discussions on tort liability are grounded in a philosophical understanding of responsibility and causation.

This book seeks to advance interdisciplinary legal scholarship in answer to the conundrums posed by this new entity in our midst. Drawing upon both contemporary and classical legal and philosophical analysis, we attempt to develop a prescriptive legal theory to guide our interactions with artificial agents, whether as users or operators entering contracts, acquiring knowledge or causing harm through agents, or as per-

sons to whom agents are capable of causing harm in their own right. We seek to apply and extend existing legal and philosophical theories of agency, knowledge attribution, liability, and personhood to the many roles artificial agents can be expected to play and the legal challenges they will pose while so doing. We emphasize legal continuity, while seeking to refocus on deep existing questions in legal theory.³

The artificial agent is here to stay; our task is to accommodate it in a manner that does justice to our interests and its abilities.