

Slaves, Vending Machines, and Bots. Samir Chopra and Laurence F. White. *A Legal Theory for Autonomous Artificial Agents*. Ann Arbor, MI: University of Michigan Press, 2011. 264 pp. ISBN 9780472051458. \$35 pbk.

Reviewed by Jorge Martins Rosa

Legal doctrine seems to be on the rise in the field of cyberspace, from the more focused Internet studies to broad-ranging cultural issues. Lawrence Lessig brought it into the limelight with his 1999 book *Code and Other Laws of Cyberspace*, in which the mantra-like “Code is law” served his claim of a structural similarity between the legal (“East Coast”) and the programming (“West Coast”) codes. Almost right after, the discussion, while becoming less philosophical and more pragmatic, widened to mundane political and economic matters such as those concerning copyright and patents, privacy, and Internet neutrality.

With Samir Chopra and Laurence F. White’s *A Legal Theory for Autonomous Artificial Agents*, although the authors rarely abstain from practical and everyday illustrations, theory strikes back. And, as we will see, in spite of following a very different route from the established debate concerning the idea of the post-human, namely because more focused on the “humanization” of machines than on the “cyborgization” of our biological self, they arrive at very similar conclusions. Being an occasionally dense book on legal doctrine—Chopra teaches philosophy at CUNY and White is a lawyer specialized in cases involving technological matters, and that dual contribution is occasionally patent, though never jeopardizing the consistency of their arguments—even the nonspecialist reader will be quickly engaged by the contemporaneity of the theme and by the vividness of the language. And it is indeed contemporary: in an increasingly automatized world where artificial entities power e-commerce websites or negotiate in stocks and bonds almost without human intervention, may the legal building still consider them mere tools? *A Legal Theory* insists on the need for a substantive reform of their status; actually, acknowledging them as full-fledged agents is the first step of “the theory that underpins this book: an agency law approach to artificial agents is cogent, viable, and doctrinally satisfying” (23).

This means that, according to Chopra and White’s claim, there is an urgency to concede them legal personality: “The most radical (and perhaps in the not-too-distant future, the most just) solution to the contracting problem would be to treat artificial agents as legal agents who have legal personality and contracting capacities in their own right” (42). Though contested (and they take their time to confront opposing arguments), it does not come as

such a radical solution if we follow their predominantly pragmatic reasoning, which starts with the basic notion of legal agency in all kinds of contracts. The concept, summed-up by the Latin *dictum* “Qui facit per alium facit per se” (17)—an agent is someone acting on someone else’s behalf, as if she were acting herself—is, as the book unravels, dissected through a closer analysis to the contracting problem (chapter 2), to knowledge attribution (chapter 3), and to tort liability (chapter 4).

Chapter 2 presents the reader with the basic ontology (and several variations) of a contract mediated by agents. A typical case, which helps us (even readers unfamiliar with legal jargon) to understand the involved parties and their roles, can be something as widespread as buying a book through Amazon (cf. 47): the *agent* (the software in the server) acts on behalf of the *principal* (Amazon, the corporation that sells books and other items), reaching and negotiating with the *third party* (the user buying commodities). But to that analogy we need to add some other features attributable to the agents, such as the recommendation engine that enhances the age-old experience of buying an item; at the same time, it points to their increasing autonomy and renders them problematic as mere tools, “since in many cases one party will be unaware of the terms of the particular contract entered into by its artificial agent” (30). Just like in the famous *boutade* by Heidegger, both parties (as well as the legal system) only notice there is an ontological conundrum when something fails: an item delivered to the wrong address, sold at a cost substantially above or below its price, or some similar glitch. Who is liable? The most common solutions, wrapping each particular exchange within an “umbrella” contract—cf. the notoriously never-read End-User License Agreements—or keeping the offer open until an explicit confirmation is given by the user, cannot be applied to all scenarios, and even less “as the degree of operational autonomy of a given artificial agent increases” (33).

That calls as much for a review of contemporary doctrine as for an inquiry in search of applicable cases in former jurisprudence. A recurrent example is the similarity between the proposed treatment of artificial agents and the way Roman law regarded slaves: able to enter into contracts on behalf of their masters, but not to sue in their own name. In an enlightening passage, the authors remind us: “The comparison of technologies of automation to slave labor is not new. Norbert Wiener famously noted that ‘the automatic machine, whatever we may think of any feelings it may have or may not have, is the precise economic equivalent of slave labor’” (41). They miss what may have been another useful illustration, one that immediately would strike sf

fans' and scholars' minds, Asimov's well-known Laws of Robotics, particularly when the increased autonomy of the agents may be enough to breach any of those laws, that is, when a "stage of development is reached where artificial agents routinely engage in deliberate misbehavior" (55).

With or without deliberate misbehavior (concerning strict liability) at the present time, the cases of tort liability—negligence and all other instances of "breach of a duty to avoid harm to others" (119)—must also be addressed. That is the main heading in chapter 4 (I will come back to chapter 3 in a moment). In this specific field of law, there is already a wide corpus of doctrine in which agency has to be attributed to an array of entities that cannot have the full status of legal persons because they are incapable of adult reasoning (e.g., children and dangerous animals) or that, being inanimate property, completely lack that capacity (e.g., chemical substances, or physical objects such as ships). The transposition to artificial agents actually seems easier in these cases: "[P]ossible sources for a theory of tort liability flexible enough to handle artificial agents include existing doctrines relating to liability for wild and domestic animals, children, unpredictable actors under supervision such as prisoners and even slaves, and ultrahazardous activities. [...] [P]roducts themselves became the source of strict liability via analogies made to dangerous animals" (120). And again, although the authors start by considering the (for the time-being more common) cases in which artificial agents can be regarded as mere tools or instrumentalities (trespass and negligence), the arguments are more acute when they must be considered agents in their own right, even if under a duty to a principal, for example in an outbreak of malware used for research purposes, or, in a plausible scenario that not long ago could only be found in sf, software designed to be self-modifying: "In that case, it will appear harder to trace the origin of the defect back to the programmer or manufacturer; an argument could be made the original design had been altered by the agent itself" (137).

Just like in the cases of strict and tort liability, attribution of knowledge also plays an important role in contemporary law, and artificial agents are far from being exempt. Actually, if we look at the most notable illustrations in chapter 3, the privacy worries raised by the algorithm behind Google's AdSense and the alleged responsibility of ISPs in cases of copyright infringement through their infrastructures, knowledge never was as much a case of concern as today. Speculating about "how a wholly physical system like an artificial agent could possess a mental state approximating belief" (74) is, again, no longer sf, although it is easy to stumble upon some claims that evoke the legacy of the genre. Fredric Brown's classic short story "Answer"

immediately comes to mind at a point when we read that attributing knowledge to artificial agents entails that the belief corpus should be defined, just as with humans, as “the set of propositions the agent is committed to, that is, those propositions p the agent answers ‘Yes’ to when asked, ‘Do you believe p ?’” (77).

These chapters prepare the ground for a thesis only fully developed in the remaining one: the legal system has already started to accommodate the fact that “things” and “persons” are not mutually exclusive categories. Bringing Kant and especially Locke to their side (cf. 171–72), the authors remind us that “personhood”—from the Latin *persona*, a mask that enables us to play a role (legal, in this case)—is a useful fiction instead of a natural attribute; as such, it should reflect the ever-changing features of each society. Even intentionality may become a secondary attribute, as Oliver Wendell Holmes, Jr. proposed: “Holmes believed [...] that the role of mental entities in law, such as ‘intention,’ should diminish as law becomes more sophisticated, [...] more a matter of conduct than of intent” (146, qtd. from Richard Posner). Or, again uncannily coming close to an infringement of Asimov’s Laws, the concept of intention may be kept as long as it is reconfigured through a purely pragmatic perspective: “[A]ll that is needed to show intent is that the agent in question picked one course of action when another, [...] also attractive, but not likely to cause harm, was not chosen” (146). Thus the hardest barrier to overcome is not the technical one, as we seem to be each day closer to having agents sophisticated enough to fulfill the conditions proposed by the authors, but rather our own stubbornness in believing we are the sole entities gifted with *free will*, *autonomy*, and a sense of *moral responsibility*. For each of these attributes, suitable and plausible arguments are evoked, always under the assumption that a pragmatic stance, not an ontological one, is the rule in legal nomenclature. The argumentation only becomes shaky when discussing a fourth (and the most pragmatic) attribute: *identification*, a hindrance the authors acknowledge, although also claiming it can be surmounted. “It is not clear whether the subject agent is the hardware, the software, or some combination of the two. To make things worse, the hardware and software may be dispersed over several sites and maintained by different individuals. [...] [U]nlimited copies of the agent can be made at a very low cost. Perhaps each instance [...] could be a separate person [...] Or consider an agent system, consisting of multiple copies of the same program in communication, which might alternately be seen as one entity and a group of entities” (181).

In spite of those details, jurisprudence—mostly focused in the US reality, but with several calls of attention to the British and European doctrines, and

also the Australian (which, interestingly, is the more closely aligned with the arguments in the book)—can give us enough examples that justify conceding artificial agents at least the status of *dependent* legal persons “for reasons of expedience, while ascriptions of full moral personhood, *independent* legal personality, and responsibility might await the attainment of more sophisticated capacities on their part” (189, emphasis added).

Their conclusion, stressing, as was announced above, that even if we remain human we are no longer alone—“In deciding that artificial agents are persons, courts or legislatures would send a message about their commonality with us” (85), and “At best it would be a chauvinistic preservation of a special status for biological creatures like us” (191)—may thus be seen as yet one more discourse heralding a post-human era. Norbert Wiener, who in *The Human Use of Human Beings* called for a redefinition of “life” in very similar terms, must be laughing in his grave. As such, the book’s appeal goes far beyond the ultra-specialized field of law practice and doctrine. Other scholars and students in humanities and social sciences will find here a rich but barely explored lode, adding to the inherently transdisciplinary nature of cybercultural studies. Even in undergraduate classes on sf, the more empirically oriented chapters may be used as the theoretical background to discuss the trope of artificial intelligence, from the Golden Age to cyberpunk.

The Alien Tooth. J. G. Ballard. *Miracles of Life: Shanghai to Shepperton, An Autobiography*. New York: Liveright Publishing, 2013. 250 pp. ISBN 9780871404206. \$25.95 hc.

Reviewed by D. Harlan Wilson

J. G. Ballard’s autobiography was published shortly before his death in 2009. This reprint is the first American edition and includes an introduction by China Miéville. Part of the point, perhaps, is to syphon the momentum of Miéville’s celebrity, which, since he won the Arthur C. Clarke Award and British Fantasy Award in 2001 for *Perdido Street Station*, has experienced a steady ascent, culminating in the Hugo he won in 2010 for *The City & the City*. It’s something of an irony that Ballard, despite being a revolutionary sf author and technician of prose, never received a major sf award and was frequently shunned in sf circles for breaking narrative rules and attempting to reinscribe the genre (at least during the first half of his career, after which he abandoned the genre, although I would argue he outgrew it). At the same time, Miéville to some degree writes out of the avant-tradition of New