ABSTRACT

In this chapter we set forth a case study of the integration of philosophy and computer science using artificial agents, beings ruled by abductive logic and emergent behavior. Our first step in this chapter is to highlight different models that we developed of such agents (a set of them related with evolutionary game theory and one model of a narrative storyteller robot). As we indicate, each model exemplifies different aspects of the bottom of the hill of autonomy as an emergent property of artificial systems specified through three aspects (“Self control”, “Adaptivity to the environment” and “Response to environment”). In summary, our conception is that autonomy, when presented as an emergent characteristic, could fill the important place given it by elaborations in philosophical ethics and one that leads us to a clearer comprehension of where to direct our efforts in the field of artificial agents. We conclude this chapter with the notion that this reevaluation of autonomy is necessary for the enhanced comprehension of human morality.

INTRODUCTION

Autonomy plays a central role in modern ethics, as on its basis we distinguish the class of ethical agents from the other beings - “only things” or at most Kant’s “heteronomous” beings, but not fully ethical agents. Forged from medieval Christian conceptions of the divine origin of morality and tempered with the rediscovery of classical Greek philosophers, with their natural order establishing standards for fitness,
excellence and virtue, traditional conceptions of autonomy continue to inform notions that often appear more central to the field. Freedom, personhood, responsibility and intentionality, aspects of autonomous agency by way of which some beings differentiate themselves from the mass of objects, tools, artifacts and things littering this physical world, at root derive their value in the aid they render in answering this fundamental question: autonomous, or not?

This binary is complicated by apparent grey-area cases. The existence of human children, incapacitated humans (mentally ill, intoxicated) and intelligent animals imply that “ethics” is not the simple domain of pure rational agents. Sensitivity to such cases is evident in the “moral patient” of Regan (1986) and Singer (1993). But, this sensitiveness remains exceptional. Autonomy often remains presented as an especial characteristic distinct to human beings differentiating them from other beings.

Our aim in this chapter is to suggest that artificial agency emerges with the multi-pole development of different characteristics, and that the development of capacities of artificial agents accordingly seeds the grey areas of agency. Ours is a proposal for the reevaluation of autonomy on the basis that we good reasons to recognize as autonomous suitably developed artificial systems. Our strategy is to open the space between programming efforts and philosophical investigation presenting, in a sense, a philosophical interpretation of a programming effort. This will allow us to redefine autonomy in a way that may prove useful. The first step is the recognition of the legitimacy in the conception of autonomy as is develops in response to challenges coming from different areas of historical investigation such as sociology and psychology. The second step, the one that distinguishes our proposal, is to present different programming efforts and to try to determine any sense of autonomy in these systems without diminishing what stands out from the preceding review of the modern tradition. This leads to our third step, the elaboration of the aforementioned frame, wherein different beings may be inserted as in a map revealing the different dimensions of autonomy as an emergent property. Lastly, in the final part of the chapter, we present some far-reaching consequences that we believe need to be addressed, and toward which we plan to aim in future investigations.

1. SETTING THE STAGE

Philosophers, particularly those in the rationalist tradition, have pictured autonomy as affording an important role not just in the practice of morality but also in the distinction of agents that are moral beings from others which are not. This tradition of so establishing autonomy or self-governance at the center of Ethics has a history that can be characterized globally and episodically. Globally, following Schneewind (1998), we may take it as a reaction against “the conception of morality as obedience” to some divine order. It is an internalization. Episodically, we may highlight the establishment of the Kantian formulation crowning autonomy as the central concept essential for morality, and thus Ethics, in the first place.

This crowning follows a natural development inside the modern moral tradition binding Ethics with human psychology. In this vein, when Kant established autonomy as the foundation of human dignity, he established an association between these two fields, something that could be verified in the following supportive statement: “Autonomy is thus the ground of the dignity of the human and of every rational nature” (Kant, 2002, AK 4:436). Autonomy reveals the particular constitution of our souls. In this sense modern Ethics, and with it the question of how we should live, merges a conception of what we should do with a conception of how our minds operate optimally.