Cartography of the Values Involved in Robotics

Daniel López, Aníbal Monasterio, Mario Toboso, Manuel Aparicio, Txetxu Ausín and Ricardo Morte

Abstract— **In this contribution, we analyse important questions related to the conceptual foundations of axiology, or the discipline concerned with the study of value, with regard to its repercussion in the field of *inclusive robotics*. We take a pluralistic position on the subject, with the goal of presenting a way for us to approach the complex system of values around which current scientific-technological practice takes place. Based on these considerations, we provide a set of ideas that can be useful as a bridge to move from the axiology of the technology towards the field of *inclusive robotics* within the paradigm of Responsible Research and Innovation (RRI).**

I. Introduction

THE idea of value and its ontology, that is, what it is and what its defining characteristics are, has been a question of interest for philosophers throughout the history of ideas. In philosophy, the expressions "axiology" or "value theory" are used with different meanings to refer to areas of scientific practice and knowledge that have an evaluative component. We can say that in addition to the traditional position, which concentrates on ethics and aesthetics, other philosophical currents such as feminism and political philosophy maintain a certain axiological perspective or contemplate a value theory at the base of their foundations. Axiology or value theory is dedicated to reflecting on what is valuable and the reasons for it being deemed so. A perennial question within value theory or axiology is to determine if values are subjective or the state of things in the world. Other traditional questions are the typology of values and how they can be analysed and decomposed into categories: intrinsic, extrinsic, instrumental values, etc., how they can be classified, if they are from a single source or if many values exist (monism or pluralism), etc.

A more interesting philosophical debate for the object of our study, and which occupied a good part of the twentieth century, was raised on the separation between science and values. For example, Bertrand Russell [1] emphatically affirmed that values are wholly outside of the domain of knowledge. As such, when we say that something has value, we simply express our emotions. In other words, if two people differ on a question of values there is no type of truth behind this, but rather a simple difference of taste. On the other hand, Kuhn [2] raised the debate again by affirming the existence of a pluralist scientific ethos that consisted of at least five universal values: accuracy, consistency, broad scope, simplicity and fruitfulness. Moreover, in a position contrary to Bertrand Russell, Hilary Putnam [3] has affirmed the objectivity of some values, including ethical ones, criticizing the fact-value dichotomy. These considerations broaden the notion of value and downplay the split between axiology and empirical science [4]. Although delving deeper into these positions is outside the scope of this communication, we can assume that scientific-technological processes, and as such robotics, have a significant axiological component.

II. Material and methods

Following are the methodological and conceptual bases of the reflections presented in this article. First, the axiology of Javier Echeverría is presented as a theoretical frame of reference to work with the complex system of values underlying scientific-technological activity. In line with this perspective, examples are given of controversies or tensions between different subsystems of values that are especially relevant for interactive robotics. Second, a methodology is proposed to move towards inclusive robotics within the paradigm of Responsible Research and Innovation (RRI). With this objective, and starting from the classic text from Langdon Winner "Do Artifacts Have Politics?", a particular reading of what is inclusive is explored and a series of criteria for evaluation (values) oriented towards this objective are proposed.

III. Results

Javier Echeverría [4] defends that scientific and technological practice are loaded with values, given that human actions themselves are guided by value systems. Minimally these values would be epistemic or cognitive (clarity, consistency, originality, verifiability, truth…) and technical (utility, innovation, trustworthiness, efficiency, accuracy…), although as we will see, other values, mainly economic, also have an influence. In Science and Values [5], Echeverría lays the foundation for his axiological proposal, of which his way of understanding values is of special interest to us. Later [6], by doing a reading of scientific-

This communication was funded by the European projects INBOTS and EXTEND of the H2020 Programme. They constitute the IFS-CSIC Research Group in the INBOTS and EXTEND projects.

D. López is a Predoctoral Researcher at the Department of Theoretical and Practical Philosophy in the Institute of Philosophy of the Spanish National Research Council, Madrid, Spain (daniel.lopez@csic.es).

A. Monasterio is a Basque Government Postdoctoral Researcher, Bilbao, Spain (anibal.monasterio@ehu.eus).

M. Toboso is a Tenured Scientist at the Department of Science, Technology and Society in the Institute of Philosophy of the Spanish National Research Council, Madrid, Spain (mario.toboso@csic.es).

M. Aparicio is an Associate Professor at the University of Murcia, Murcia, Spain (manuel.aparicio@um.es).

T. Ausín is a Tenured Scientist at the Department of Theoretical and Practical Philosophy in the Institute of Philosophy of the Spanish National Research Council, Madrid, Spain (txetxu.ausin@cchs.csic.es).

R. Morte is a Ph.D. Student at the University of Granada, Granada, Spain (ricardo63@autistici.org).

technological practice throughout the twentieth century, the term *technoscience* is coined to refer to projects characterized by a close connection between engineers, technicians, scientists, politicians and even military institutions (for example, information technologies, space exploration, the Genome Project). From this structural analysis, what interests us is the pluralism of values involved in *technoscientific* practice, as well as the conflicts over values between different agents. Which elements from this axiological approach are especially relevant for interactive robotics?

First, from the start we forego a definition of values, but will instead explain some of their characteristics. Continuing with Echeverría [5], an axiological term does not have meaning on its own. On the contrary, it only acquires meaning when, inserted into a set of values, an agent applies the axiological term in question to a system or thing, creating a valuation. In this way we go from logic based on subjects and predicates (for example, saying that an agent is innovative) to logic based on arguments and functions (certain actions of an agent have been innovative). This allows specific cases of innovative actions to be analysed as well as different degrees of innovation to be compared. As such, graduality of values is acknowledged. In other words, a value, for example, *utility*, can be satisfying to varying degrees. Also, it is recognised that evaluations are actions with a plurality of components that can be nuclear (if they need a minimal degree of satisfaction to be accepted) or orbital (if they do not). Finally, it is recognised that evaluations are always carried out by an agent, whether it is this person or many (for example, a group of researchers or a parliament), following a value system. Although this axiological framework has more characteristics than those outlined here, their implications remain outside the scope of this communication (see [5], [6]).

Up to now we have focused on describing a general way of approaching values. Focusing more specifically on robotics, an analysis of publications from different interest groups can be useful when it comes to drawing a map of values. Each company in the robotics industry has a series of values that guide its mission and business activity. We have taken as a source various documents on values in the robotics industry, among them the Executive Summary World Robotics 2017 Industrial Robots. In this case, we have identified: innovation, security, profit maximisation, excellence, respect. The values we took from the users as stakeholders were obtained from a qualitative review of various publications from ASPAYM (Association for People with Spinal Cord Injury). For them, the values that stand out are quality, convenience, usefulness, functionality and price. The values from society are, on the whole, generic but are present in numerous reports on Corporate Social Responsibility. In this case, we have identified community, responsibility, trust, integrity, inclusivity and diversity. As a representational method for values we use spatial and relational distributions in the form of a diagram, in this way determining which values guide these three interest groups, that is: the robotics industry, users and society as a whole.

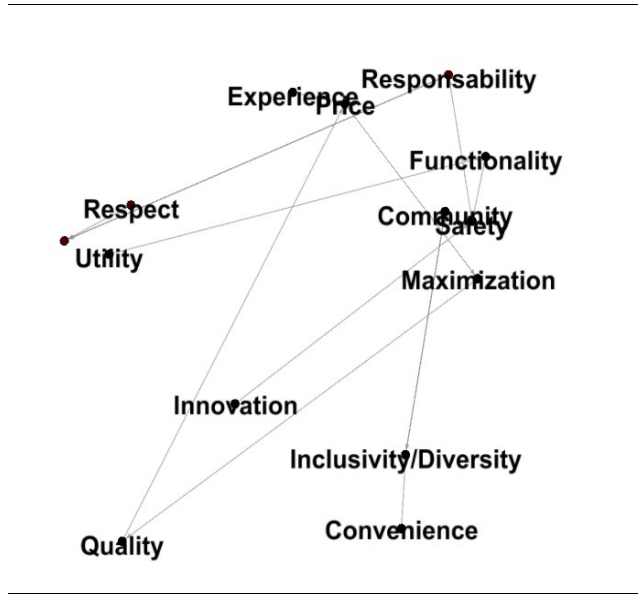

Figure. 1. Network of value nodes from the three stakeholders (robotics industry, users, society). Representation of how the values connect with each other. Created with Gephi 0.9.2.

In the same way that the concepts of robot, machine and technology are blurring and overlapping [Conceptual Analysis: technology, machine and robot, in press], robotic technologies are also blurring the borders between human subjects and technological objects. This blurring of lines has ethical implications affecting our axiological categories. In this sense, values traditionally reserved for human beings (such as autonomy, responsibility, creativity) are beginning to be projected onto intelligent technological entities, often driven by our tendency towards anthropomorphism when it comes to contemplating and evaluating robotics.

Advances and developments in robotics and Artificial Intelligence (AI) are making us aware of the numerous ethical implications of these technologies on different groups of people.

In the case of inclusive robotics, it is not the same how children will be affected as compared to adults, and even gender bias must be considered for inclusive robotics having to do with care that take into account the needs and demands of individuals with disparate personal characteristics.

We believe that to identify the values present in the design of robotics, it is necessary not only to take into account all interest groups, but also that preferences, axiological hierarchies and how they come into play in different people should be tested empirically.

A recent study [7] called the Moral Machine experiment is a global analysis that attempts to quantify social expectations on the ethical principles that should guide the behaviour of machines. More than two million people have participated in this study (and the platform continues to be open in 10 languages: http://moralmachine.mit.edu/) with nearly 40 million moral dilemmas resolved. The idea was to create an open collaborative platform to record people's decisions facing moral dilemmas: such as what should a self-driving car prioritise when it finds itself in a dilemma. Humans over pets? passengers over pedestrians? more lives over fewer? The data produced can offer ideas on the collective ethical priorities of different cultures.

The experiment concerns self-driving cars, but many of its conclusions can be extrapolated to other types of robots.

IV. Discussion

Technologies do not only transform objects, but also transform habits, customs or relationships [6]. Consequently, if we understand robotics systemically, we can say that it will transform systems, whether these are social, economic or natural. This transformative vocation creates conflicts of values in multiple phases of technological development. For example, when it comes to evaluating which technologies get financed, controversies arise over different priorities. Later, given that innovations compete in the market, this becomes a very important evaluation mechanism. When innovations reach society, controversies are often posed in terms of technophilia and technophobia. On these value conflicts underlie moral, social, religious, aesthetic and ecological values, to name a few. In this sense, the context where robotics are applied determines, delimits or complicates the axiological pluralism we make reference to (for example, creating controversies with other subsystems of values in the fields of care, health or the military). Even so, given that interactive robotics includes all robots that maintain physical, cognitive and emotional interaction with a person, some controversies inherent in the technology itself can be identified. For example, *utility* or *efficiency* in these robots' performance requires deep understanding and interpretation of the movements, thoughts and emotions of the people they interact with. Economic and technical values may conflict with social and moral values such as *autonomy* and *privacy*.

Nevertheless, normatively, what can this axiological model contribute to an inclusive robotics? In the first place, to empirically study how other systems doing inclusive actions are evaluated. In this sense, a good starting point would be *inclusive education*. Second, to defend the consideration of this value as a central value that requires its degree of satisfaction to be equal or greater than a determined level or minimum benchmark [5]. Third, to locate and analyse those evaluation mechanisms that take place throughout the entire innovation process of interactive robotics. Finally, to propose *inclusion* as a central value within these mechanisms as well as giving more positive weight to other values related to it. This process will take us in the same direction as the set of ideas and initiatives brought together in the paradigm of Responsible Research and Innovation (RRI) [8].

That said, we consider that this approach should be complemented with practical analysis of the interactive robots that come out on the market. Although the theoretical approach to technological innovations is an extremely important question, auditing the political properties of these robots in their interaction with human beings should be a significant objective. Langdon Winner argued convincingly that artifacts can have political characteristics in two distinct ways [9]. First, when their design, invention and implementation become a means to reach an end. The example of Moses's bridges used by Winner, although it can be classified as a conspiracy, points to how technologies are ways of ordering our world that, at times, can be designed with the intention of producing a series of logical consequences from their implementation. In this case, Moses's public works (motorways and bridges) were designed with the objective of favouring the use of automobiles, limiting access by public transportation to Long Island beaches. As such, only the rich and affluent classes could freely use the beaches for their fun and leisure activities. Secondly, Winner points out that there are certain technologies whose own nature is very specifically politically charged. As such, adoption of a certain technological system would imply a series of determinants regarding human relations, favouring political and moral values such as centralisation or decentralisation, equality or inequality, repression or liberation. In his article, he uses the example of solar energy as a technological system favourable to a democratic or egalitarian society, in contraposition to other types of technologies based on nuclear energy, coal or oil that require a higher degree of centralisation. In other words, a technology should not be evaluated only on how it contributes to efficiency and productivity, but also for the way in which it can create certain forms of power and authority. As such, these political characteristics can work in favour or against *inclusivity*.

This perspective allows us to explore another road towards an inclusive robotics. Part of society looks with concern at the dangers of a technified society. This feeling has to do with the alienating nature of many technical systems, in the sense that the goals and values of who is interacting with the system are often seen as being subject to the artifact [10]. They influence and condition our life, but we practically cannot influence them. Faced with this situation, Miguel Ángel Quintanilla [10] coined the term "engaging technologies," as opposed to other alienating technologies, proposing a series of evaluative criteria for development based on these technologies. Next, we explain and adapt the criteria we consider to be the most relevant for an inclusive robotics aligned with the paradigm of Responsible Research and Innovation (RRI). *Openness*: a robot is open if its software is free. *Versatility*: a robot is inclusive if it allows alternative uses by its operators or users. *Docility*: a robot is docile to the extent that its functioning, control and shutdown easily depend on a human operator. *Obsolescence*: a robot is inclusive if its repair, updating and maintenance are promoted above its substitution. *Comprehensibility*: a robot is understandable if it has some instructions and a comprehensible design for its operators or users.

V. Conclusions

Technologies are the result of a plurality of agents who take decisions to make them how they are. These decisions involve a diversity of values, among which the values that are technical, political and economic often prevail. As robotic technology becomes more autonomous, it is necessary to identify the values and ethical principles that should regulate the interaction of robotic systems with human beings. The fundamental challenge for inclusive robotics is the creation of a global behavioural strategy for interactive technologies, similar to Asimov's Three Laws of Robotics, but with empirical validations and recognition of the global plurality of values. These ethical principles and values should include: a) human dignity, b) value of autonomy, c) value of privacy, d) principle of nonmaleficence, e) principle of responsibility, f) principle of beneficence, and g) value of justice. As RRI points out, inclusive and participatory methodologies should be used in all the stages during the research and innovation processes [7]. This movement implies "bringing social and moral values" from the citizenry to processes that traditionally have acted independently, something that, as we have seen, does not necessarily mean neutrality. This profoundly complex scenario is what motivates our axiological proposal, aimed at making it easier to satisfy desirable values and criteria such as *inclusivity,* putting the different disciplines that form part of the development of robotics into the same conceptual framework. A pluralistic approach with these characteristics could lessen the risks and make the most of the opportunities from innovations in robotics, including the diversity of agents and values that will be involved in its development in the decision-making process.

References

[1] B. Russell, *Religion and Science*. Oxford University Press, 1935.

[2] Kuhn, *The Essential Tension*. University of Chicago Press, 1977.

[3] H. Putnam, "Beyond the Fact–Value Dichotomy," *Critica*, vol. 14, no. 41, pp. 3–12, 1982.

[4] J. Echeverría, "Ciencia y valores: Propuestas para una axionomía de la ciencia," *Contrastes. Revista Internacional de Filosofía*, 1998.

[5] J. Echeverría*, Ciencia y valores*. Destino Barcelona, 2002.

[6] J. Echeverría, *La revolución tecnocientífica*. Fondo de Cultura Económica de España, 2003.

[7] E. Awad, S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff, J. Bonnefon and I. Rahwan, "The Moral Machine experiment", *Nature*, vol. 563, no. 7729, pp. 59-64, 2018.

[8] "About RRI - RRI Tools." [Online]. Available: https://www.rri-tools.eu/about-rri

[9] L. Winner, "Do Artifacts Have Politics?" *Daedalus*, vol. 109, no. 1, pp. 121–136, 1980.

[10] M. Á. Quintanilla Fisac, "Engaging Technologies: Criteria for an Alternative Model of Technological Development," in *Spanish Philosophy of Technology: Contemporary Work from the Spanish Speaking Community*, B. Laspra and J. A. López Cerezo, Eds. Cham: Springer International Publishing, 2018, pp. 103–123.