https://doi.org/10.17163/soph.n30.2021.07

Didactical projections of the argumentative theory of reasoning

Proyecciones didácticas de la teoría argumentativa de la razón

Rodrigo Sebastián Braicovich[*]
Independent researcher, CONICET, Rosario, Argentina
rbraicovich@gmail.com
Orcid code: https://orcid.org/0000-0003-2785-7293

Abstract

The goal of the article aims at establishing a dialogue between three lines of inquiry within contemporary epistemology: Virtue Epistemology, Bounded Rationality and Argumentative Theory of Reasoning. Faced with the problem that we are interested in dealing with here, i.e., the search for a theoretical framework that might allow us to design pedagogic strategies (both within the framework of the didactic of philosophy and outside of it) based on realistic premises, Virtue Epistemology will be presented here as a strongly optimistic current from an epistemic viewpoint. The paradigm of Bounded Rationality will represent the exact counterpart, insofar as it seems to lead to a pronounced pessimism concerning the possibility of designing strategies that may allow us to improve the agent's epistemic practices. In the middle of these two extremes, the Argumentative Theory of Reasoning (developed in the last decade by Hugo Mercier and Dan Sperber) represents a promising alternative for two reasons: in the first place, because it offers an answer to the problem (faced by the paradigm of Bounded Rationality) of the adaptive character of human reason from an evolutionary viewpoint; secondly, because it allows us to overcome the epistemic pessimism that is essential to the paradigm of Bounded Rationality when planning pedagogical strategies that are not only realistic but also effective.

Keywords

Argumentation, rationality, critic, bounds, virtue, epistemology, evolution.

Suggested citation: Braicovich, Rodrigo (2021). Didactical projections of the argumentative theory of reasoning. *Sophia, colección de Filosofía de la Educación, 30*, pp. 197-215.

[*] Doctor in Humanities and Arts with a mention in Philosophy. He has published articles on cognitivism and psychology of action from the perspective of classical antiquity and from contemporary thought. This research has the support and financial support of the National Council for Scientific and Technical Research (CONICET) of Rosario, Santa Fe, Argentina.

Resumen

El objetivo del artículo consiste en poner en diálogo tres líneas de investigación dentro de la epistemología contemporánea: la Epistemología de la Virtud, el paradigma de la Racionalidad Limitada y la Teoría Argumentativa de la Razón. Frente al problema que interesa analizar aquí, a saber, la búsqueda de un marco teórico que permita diseñar estrategias pedagógicas (tanto al interior de la enseñanza de la filosofía como fuera de ella) sobre premisas realistas, la Epistemología de la Virtud será presentada como una corriente marcadamente optimista desde el punto de vista epistémico. El paradigma de la Racionalidad Limitada representará la contrapartida de dicha corriente, en la medida en que parece conducir a un pesimismo marcado respecto de la posibilidad de diseñar estrategias que permitan perfeccionar las prácticas epistémicas de los sujetos. Frente a estos dos polos, se sugerirá que la Teoría Argumentativa de la Razón (desarrollada en la última década por Hugo Mercier y Dan Sperber) representa una alternativa prometedora por dos razones fundamentales: en primer lugar, porque ofrece una respuesta al problema (enfrentado por el paradigma de la Racionalidad Limitada) del carácter adaptativo de la razón humana desde un punto de vista evolutivo; en segundo lugar, porque permite superar el pesimismo epistémico esencial al paradigma de la Racionalidad Limitada al momento de planificar estrategias pedagógicas realistas y efectivas.

Palabras clave

Argumentación, racionalidad, crítica, limitada, epistemología, virtud, evolución.

Introduction

The cognitive revolution that took place in the middle and end of the last century forced a substantial revision of the classical conception of human rationality, that is, of the conception of it that the West had inherited from classical thought and which had found its paradigmatic expression in regards to epistemological optimism, in Enlightenment thought. The paradigm of Bounded Rationality, in particular, contributed to a decisive questioning of the traditional trust in the essential perfectibility of the human species from a rational development (both individual and collective) that would operate as the key to human progress. Herbert Simon's initial studies of the epistemic limits inherent in human decision processes, together with the systematic study of cognitive biases, led, in effect, to an essentially bleak picture of trust in the human capacity to instantiate efficient and rational decision-making mechanisms. Faced with such a scenario, the Argumentative Theory of Reasoning, developed in the last decade by Hugo Mercier and Dan Sperber, appears as an innovative alternative, insofar as it allows a rereading of the supposed limitations of rationality from reinsertion of human cognitive architecture in its original evolutionary setting. As we will try to show, even though this approach does not endorse the reinstatement of the exacerbated optimism in human rationality typical of the classical conception, it allows at least to escape the markedly pessimistic paths of the Bounded Rationality pa-

Print ISSN: 1390-3861 / Electronic ISSN: 1390-8626, pp. 197-215.

radigm, while opening avenues of exploration and design of realistic and effective teaching strategies from the point of view of argumentation.

The exhibition will be articulated in the following manner: in the first section, the optimism that essentially and explicitly characterizes the most important and developed aspect of the Virtue Epistemology will be exposed, showing how it represents a continuity with respect to the classic paradigm of rationality (both ancient and Enlightened). The second section will address the Bounded Rationality paradigm as an exact counterpoint to the optimism of the classical paradigm, insofar as it aims to highlight the structural limitations of human cognitive architecture, either through the study of limits in the computability of the problems faced by human beings in decision-making contexts, such as through research on the cognitive biases that run through human reason. Faced with both extremes, the third section will try to show that the Argumentative Theory of Reasoning, even when it starts from the paradigm of Bounded Rationality, offers a novel alternative to such extremes, insofar as it allows us to understand the emergence of human Reason in light of an evolutionary landscape that explains the central role that persuasion assumes in its operation. The fourth section addresses the reasons why the Argumentative Theory of Reasoning allows to reinterpret the cognitive biases studied from the paradigm of Bounded Rationality, understanding (at least) some of them not as shortcomings of human reason but as characteristics that are a positive return once they are restored to their original evolutionary stage. The last section proposes some projections of the Argumentative Theory of Reasoning in the field of pedagogy and, especially, at the time of designing didactic strategies both within the teaching of philosophy and outside it.

The epistemic optimism of classical rationality and the Virtue Epistemology

As a distinctive epistemological current, the Virtue Epistemology has been traversed from its very beginning in the last decade of the last century by the presence of two clearly differentiated aspects with little mutual interaction. The first of these aspects is represented by the reflections of Ernst Sosa and John Greco, fundamentally, in relation to the problem of trustworthiness, and one of its fundamental objectives is to respond to the problem of skepticism[1].The second aspect that opens up, simultaneously, within this current is represented by an extremely diverse and growing set

Print ISSN: 1390-3861 / Electronic ISSN: 1390-8626, 8626, pp. 197-215.

of reflections of a normative nature in relation to the ways and strategies to generate and/or become ideal epistemic agents[2]. Such is the divergence between both aspects, that the second of them not only does not recognize Sosa and Greco (or A. Goldman) as relevant antecedents for their own explorations, but even some of their defenders consider the problem of skepticism as a second-order problem. The explicit starting point of this aspect, on the contrary, is found in the reflections initiated, in the middle of the last century, by the model of the Ethics of Virtue, and, based on this, it tends to interpret the epistemic virtues more as character traits than as instruments to reach the truth (as Sosa and Greco did)[3].

But what are the epistemic virtues and what differences them form the ethical or moral virtues? In general, enough terms to cover the broad spectrum of virtues analyzed by the authors (and resorting to a formulation of Aristotelian inspiration), epistemic virtues can be defined as stable dispositions of character that allow the subject to carry out practices related to knowledge in an ideal form, or, alternatively, the dispositions of the character necessary to carry out an honest, careful, sensitive to detail, deep, persevering, thoughtful, cautious investigation, etc. What defines, then, the epistemological virtues and the differences from the ethical virtues is the fact that they relate specifically to the problem of knowledge and not to the problem of the relationship with others (not at least directly). And this determines, in turn, the spectrum of virtues that will be the object of study by epistemologists: epistemic courage, epistemic humility, open-mindedness, epistemic rigor, etc.

The classical roots of this conception of what should be an optimal exercise of rationality are evident not only in terms of the idealized and optimistic conception that epistemologists of virtue possess of human rational capacities but of the very concept of epistemic virtues. —Concept that is modeled on the Aristotelian treatment of the ethical and dianoetic virtues (and that inherits, on the other hand, one of the central problems that afflict the moderate cognitivism of the Aristotelian approach, namely: the problem of *akrasia*[4]) —. This Aristotelian matrix that is at the basis of the Virtue Epistemology program has proven to be extremely fruitful in organizing research on intellectual virtues, not only in relation to the definition and conceptualization of the different possible virtues that characterize optimal rational exercise, but also in terms of detecting possible obstacles to acquiring these virtues, and in terms of reflecting on the possible ways of accessing them.

From the specific level of pedagogy, this has led to outlining specific strategies that would allow the student to achieve objectives such as

Print ISSN: 1390-3861 / Electronic ISSN: 1390-8626, pp. 197-215.

carrying out a systematic and rigorous investigation, exercising a rational and solid defense of their own conclusions even in front of an adverse audience, or undertaking a dialogue/investigation with an open mind to opinions that are completely opposite to their own, forcing themselves to respect the rules of justice, tolerance, and patience with the interlocutor. The careful and in-depth analyzes that the defenders of Virtue Epistemology have offered both of the isolated epistemic virtues and of the virtuous epistemic agents have represented, in this sense, the fundamental and most innovative contribution of this current to the field of education, insofar as they offer an integrated framework, and at the same time flexible and dynamic, to build ideal knowledge practices, embodied in an ideal of a rational, critical and self-critical knowing subject.

Since Zagzebski's first steps three decades ago, thereby opening up an absolutely new terrain of theoretical exploration, Virtue Epistemology has produced a surprisingly strong corpus of pedagogical tools designed not only to assist in the acquisition of epistemic virtues on the individual plane but, in the long term, to collaborate with what some of the most optimistic epistemologists consider a true transformation of the world based on the development of these virtues[5]. It is precisely in the essentially practical dimension that epistemology carried out from the Aretaic perspective assumes, at the same time, advocating the abandonment of abstract traditional gnoseological problems, where precisely the strongest reason for its appeal lies, both for those who are dedicated to research and for those who are dedicated to teaching or reflecting on it. Now, how realistic is the epistemic ideal defended by the Virtue Epistemology? How accessible are the epistemic virtues defended by these authors and how surmountable are the obstacles whose existence they themselves recognize? Is it, after all, a viable research program? In subsequent sections, it will be suggested that the paradigm of bounded rationality and, more specifically, the Argumentative Theory of Reasoning offer reasons to cast serious doubt on the optimism that characterizes the Virtue Epistemology regarding the natural capacities of the subject of reaching (or at least approaching) the ideal of epistemic subject defended by the said current. In spite of this, it will be suggested that the specific variant of the Argumentative Theory of Reasoning opens some paths that, against the pessimism that systematically crosses the paradigm of Bounded Rationality, allows us to think about the effective restructuring of the Virtue Epistemology that leaves behind the enlightened optimism typical of the original versions of these currents.

Sophia 30: 2021.

Print ISSN: 1390-3861 / Electronic ISSN: 1390-8626, 8626, pp. 197-215.

The epistemic pessimism of the Bounded Rationality paradigm

As a general and programmatic model, the Bounded Rationality paradigm finds its origins in Herbert Simon's (1955) early reflections on the limitations that inevitably structure human rationality. The core of this paradigm, in general terms, consists of the idea that human rationality is limited in its operations by a series of factors (time, computability of the problem, epistemic limitations, etc.) that render its performance necessarily suboptimal. Human reason does not operate, in other words, as a perfect and error-free inferential machine, but, on the contrary, is traversed by serious limitations in its operations that call into question the ideal character that the classical tradition had assigned to it. Far from representing a truism, the fact that this idea has lost part of its controversial and counter-intuitive character in the academic field, it should be noted, is fundamentally due to the impact that said paradigm had from its inception in the field of social sciences and in certain sectors of the humanistic disciplines. But it is necessary to bear in mind that the limited conception of rationality represented a decisive break with a tradition that had been virtually hegemonic from classical antiquity until the 1950s, and that started from the presupposition of the absolute rationality of the underlying inferential processes to every human decision.

The most important and systematic developments of this paradigm, however, began to come from the research carried out around cognitive biases by Daniel Kahneman and Amos Tversky in the 1970s (Kahneman & Tversky, 1979; Tversky & Kahneman, 1981), researches developed fundamentally against the *Rational Choice Theory*, which is the theory that, in the field of economics, at that time embodied the most paradigmatic expression of the rationalist assumptions of the classical tradition. What did the theory of cognitive biases propose? In general terms, that human reason is completely vitiated by cognitive biases, that is, spontaneous, unconscious, and intuitive tendencies to process information from the environment by resorting to inferences that have nothing to do with the model of rationality proper to the classical logic. The repertoire of biases studied by both authors became ever greater and deeper, integrating trends such as 'availability bias' (the tendency to take into account, in decisions, only the information that is at hand, rather than looking for the most relevant information), 'false consensus bias' (the tendency to assume that what one believes or values is more widespread in the population than it actually is), or 'confirmation bias' (the

Sophia 30: 2021.

Print ISSN: 1390-3861 / Electronic ISSN: 1390-8626, pp. 197-215.

tendency to pay attention only to information that confirms one's beliefs and to dismiss information that contradicts them).

At the end of decades of research on these trends, the conclusions reached by those who, along with Kahneman and Tversky, dedicated themselves to studying these biases, were extremely negative in relation to the human capacity to put into play a critical, self-critical, and efficient exercise of rationality, conclusions that seem to openly contradict the practical expectations that cross the reflections developed within the Virtue Epistemology, and that seem to condemn to the limbo of the pure idealistic reverie of the pedagogue the possibility of design strategies that allow the optimal development of epistemic virtues. And it is precisely as an alternative to this scenario of epistemic pessimism that the paradigm of Bounded Rationality seems to inevitably lead to, that the Argumentative Theory of Reasoning model will emerge at the beginning of the last decade.

The Argumentative Theory of Reasoning: The emergence of Reason and the evolutionary landscape

The systematic study of cognitive biases showed that the traditional rationality model (defended in classical economics, as indicated by the Theory of Rational Action) did not represent at all the real dynamics of decision-making: both when consuming goods, products, or services, such as when making decisions of a political or personal nature, or when facing situations related to couple relationships or one's professional career, the decisions that are made are, for the most part, the result of inferential processes unconscious, biased and, in most cases, logically deficient. A quick reading of some of these biases, as well as the frequency with which they are brought into play on a daily basis in decisions, cannot help but lead us to wonder, curiously, how it is possible that, as a species, it has been possible to get this far (in terms of survival) with such a poorly built vehicle.

This question, on which authors such as Stanovich and West (2000, 2008), Evans (2008), Evans and Stanovich (2013), Bargh and Chartrand (1999), among others, have written in recent decades, is precisely the one that operates as a trigger for the Argumentative Theory of Reasoning: Is it really conceivable that a capacity as deficient as this has arisen as a product of natural selection, considering the serious deficiencies attributed to it from the theory of cognitive biases and from other systems of dual processing? From the perspective of classical evolutionism, a phenotypic modification in an organism becomes the object of natural selection only

Print ISSN: 1390-3861 / Electronic ISSN: 1390-8626, 8626, pp. 197-215.

if it is adaptive, that is: if it confers on said organism (and its offspring) an adaptive advantage in a certain evolutionary scenario. If it is admitted, as both Simon and Kahneman and Tversky do, that the theoretical model of evolution by natural selection is the best model of explanation that has been proposed so far to explain not only the emergence, mutation, and extinction of species, but also the emergence, mutation, and extinction of phenotypic traits, the image that the theory of cognitive biases portrays of human reason seems to be markedly non-adaptive. How, then, to explain its emergence and proliferation? It is there, precisely, where TAR introduces its most interesting argumentative turn in relation to the Bounded Rationality paradigm, without decoupling from it[6]: according to Mercier and Sperber (2019), human reason is markedly deficient, in effect, but only if it is assumed that the objective of reason is to reach the truth, or to achieve an increasingly adequate knowledge of the world (as they assumed was the case, from ancient times onwards, both philosophy and classical psychology).

But is that really the case? Is it really true that when in an argumentation what you are trying to do is seek the truth, get to know the world better and the situation in which you find yourself, simply to understand it or to be able to make the best possible decision? Mercier and Sperber (2019) suggest that no: when arguing, when reasons are demanded and one's own reasons are offered, what is being sought, in most cases, is not the truth, but rather to persuade the other regarding the truth of one's position. In this way, we find here one of the two central distinctions that TAR forces us to make in order to understand the operative mechanisms of human reason, namely: the distinction between natural and artificial contexts of reason's operation. This distinction is central for two reasons: first, because it defines the specific context of the emergence of reason: the dialogue with the other. The reason, in this sense, is a product designed for public consumption[7]: an isolated individual, who lived alone in the middle of the jungle and had no contact with other individuals of the same species, would never feel the need to argue in favor of his/her own beliefs, or even reflecting on the reasons that lead him/her to do what him/her does.

This brings with it a second element that the authors care to emphasize, and it has to do with the (non) place assumed by knowledge and truth in the development of rationality: in most natural contexts of argumentation (from the point of view of the evolutionary scenario), as already mentioned, the search for truth is not the objective at all; it is merely persuasion that is being pursued, generally by any means —and at almost any cost. The truth, after all, has as much survival value as the font

Print ISSN: 1390-3861 / Electronic ISSN: 1390-8626, pp. 197-215.

chosen by the publisher when buying a book. In the field of spontaneous human interaction, what prevails, at least from the perspective suggested by the authors, is not the truth, but the effects of a certain discourse on the interlocutor (s). And this is fundamentally due to the fact that the dialogue with the other in which the ability to argue emerges as an evolutionary niche is not an objective, cold and speculative dialogue, nor is it a neutral or consequence-free exchange: in the evolutionary landscape in the one that gradually takes shape human reason, convincing or not convincing the other can mean the difference between having access to certain goods, settings or situations, or not having it. Taken to an extreme, mastering that ability can mean the difference between survival or extinction. Reason is, considered from this perspective, a social and agonistic product: it is the daughter of conflict, of the struggle for access to certain goods and advantages - be they symbolic or material.

Far from being reduced to a relapse into the pragmatic horizons of classical sophistry, the perspective addressed by this theory aims to help understand human reason as a historical phenomenon, as a product marked by the evolutionary scenario in which it arose, and whose marks they are still present in its current structure. The displacement that this operates with respect to the classical conception of reason (on which much of the reflections and projections around the didactics of philosophy is based) is evident: while the traditional conception interprets reason as a tool In order to reach the truth, objective consensus, etc., TAR places reason in the natural environment of human evolution, and postulates the scenarios of objective, neutral and disciplined search for the truth as artificial or directly unnatural scenarios.

As noted above, from the evolutionary paradigm from which the authors start, a certain phenotypic trait (in this case human reason) is adaptive to the extent that it fulfills the function for which it was 'selected', and can operate in a suboptimal manner when put to work in alternative scenarios: just as the hand of a chimpanzee or a bonobo cannot be expected to be efficient in playing the clarinet, neither should human reason be expected to be efficient in the objective search for truth —simply because it is not the role for which it was selected. When human reason is removed from the horizon of agonistic argumentation in which it evolved and put to work in another setting, it is logical that its performance is poor, and it is logical that it is traversed by completely counterproductive biases. None of this implies, of course, that all the operations of reason are ineluctably guided by the search for persuasion, or that human beings are incapable of planning and sequencing solid arguments in favor of

Print ISSN: 1390-3861 / Electronic ISSN: 1390-8626, 8626, pp. 197-215.

their own beliefs. All that TAR affirms is that when the opposite happens, that is, when the subjects privilege the search for the truth over the persuasion of the interlocutor, or when they design and plan along logically structured and sequenced argumentations, they are faced with different situations from the original evolutionary scenario of reason.

The reconsideration of cognitive biases

I said at the beginning that TAR introduces a break within the Bounded Rationality paradigm, but without abandoning the general horizon defined by said paradigm. For TAR, reason is, in effect, traversed by cognitive biases that permeate its operations, and its performance is undoubtedly tied to limitations such as time or the computability of the decision alternatives. Mercier and Sperber's criticisms of the Enlightened optimism of currents such as the Virtue Epistemology are, in this sense, as strong as those of Kahneman (2012), Evans and Stanovich (2013), Gigerenzer (2008), or Nickerson (1998).

However, the consideration of the evolutionary landscape in which human reason arises leads the authors to make a reconsideration of the argumentative efficiency of reason, thereby tempering, at least in certain aspects, the epistemic pessimism typical of the Bound Rationality paradigm. But why is this reconsideration due? It was previously stated that, in its daily operations, and as this last paradigm has insisted ad nauseam, human reason is not particularly efficient when it comes to producing solid and systematic arguments, which is due, according to Mercier and Sperber (2019), to the fact that this is not precisely the function for which it was selected by the evolutionary process. The other side of this argument, however, has been virtually neglected and consists in the fact that, as suggested by a battery of experiments from experimental psychology reviewed by the authors, reason is extremely efficient when it does what it is designed to do. to do, namely: argue in an agonistic context.

The informal discussion contexts that more adequately represent the evolutionary scenario of reason represent, to some extent, the antithesis of the courts of justice, or of the logical, exhaustive and systematic argumentation scenarios that the traditional conception of rationality has used as criterion for evaluating the efficiency of human reason.

Unlike the sequenced, planned, and articulated argumentation of philosophical treatises or legal arguments, the dynamic that is spontaneously established in informal discussion contexts is essentially interac-

Print ISSN: 1390-3861 / Electronic ISSN: 1390-8626, pp. 197-215.

tive, which implies that participants exchange a succession of arguments brief and often impromptu or, at the very least, appropriate to the specific circumstances of not only that particular discussion, but also the specific moment of the discussion. In such scenarios, being "lazy" is an understandable and sensible decision, and this for two reasons. The first of these is that having extensive arguments prepared in advance for each of the statements themselves would require unsustainable cognitive work. The second reason is that it would probably represent unnecessary work, since, on the one hand, it is most likely that most of these claims will not be contested, and, on the other, because the dynamic nature of the dialogue allows new reasons to be improvised when it has been failed in trying to convince the interlocutor.

This relocation of reason in the evolutionary context allows a decisive rereading of cognitive biases that forces us to qualify some of the most pessimistic conclusions (from the epistemic point of view) reached by Kahneman, Tversky, Nickerson, and others, since it allows think that at least some of the cognitive biases studied exhaustively from the paradigm of Bounded Rationality may not be, strictly speaking, deficiencies of reason, but rather positive characteristics. Mercier and Sperber's extensive analysis of the confirmation bias is a paradigmatic example of their proposed reinterpretation of these biases and allows us to glimpse the hermeneutical advantages of the approach proposed by the authors.

Indeed, the confirmation bias (that is, the unconscious and spontaneous tendency to pay attention only to the information that confirms one's own beliefs and to dismiss those which do not), traditionally considered one of the most harmful tendencies since the classical study de Nickerson (1998), appears from the TAR perspective as a necessary characteristic when considering informal contexts of discussion and the essentially agonistic and persuasive (evolutionary) function of human reason. The reasons for this are clear: when you want to convince an interlocutor to accept your own belief as valid or true, what you need to do is find reasons that confirm your own position, and not his, and any information or argument that can undermine that persuasive goal becomes absolutely irrelevant. Considered from this perspective, then, the confirmation bias demands to be understood not as a weakness of human reason, but as a strength, to the extent that it allows the subject to actively seek and deploy the reasons that support the beliefs that try to impose. As Mercier and Sperber (2017) point out:

The biases and laziness of reason are not flaws; they are characteristics that allow reason to fulfill its function. Individuals have a tendency

Print ISSN: 1390-3861 / Electronic ISSN: 1390-8626, 8626, pp. 197-215.

(bias) to find reasons that support their own point of view because this is how they can justify their actions and convince others to share their opinions. One cannot justify oneself by presenting reasons that refute one's justification. One cannot convince another to change his mind by giving him arguments against it or in favor of the idea that he wants to make him abandon. And if people reason lazily, this is because, in typical interactions, that is the most efficient way to proceed. Rather than doing the hard work of anticipating counter-arguments, it is generally more efficient to wait for the interlocutor to do so (if at all) (p. 331).

To this is added, finally, a final decisive distinction that the authors propose to understand the spontaneous dynamics of reason in natural contexts, which is the distinction between the efficacy of reason at the time of producing arguments and its efficacy in evaluating arguments proposed by others. According to the authors, and again relying on the results of a set of research from experimental psychology, the limitations of human reason when producing arguments are not replicated when evaluating them: if at the time of the production of arguments one is lazy, superficial, etc., and is constantly crossed by biases (which, considered in itself, is not, as already indicated, a flaw, but something to be expected), when evaluating the arguments proposed by the interlocutors, it is much more effective, profound and critical. As Mercier and Sperber (2017) point out: "Individuals have the ability to reason objectively, rejecting weak arguments and accepting those that are solid, only they do not use these capacities on the reasons that they themselves offer" (p. 235) This is precisely what one would expect from a tool, such as the human capacity to argue, born in an agonistic context, in which the energy devoted to the active persuasion of the interlocutor ends up turning, when the roles are exchanged, into a defensive energy, embodied in the critical examination of the opponent's arguments.

Pedagogical projections of the Argumentative Theory of Reasoning

The central contribution of TAR in relation to the problem of epistemic optimism that characterizes currents such as the Virtue Epistemology, and with the epistemic pessimism that characterizes the Bounded Rationality paradigm, consists, as I have tried to show, in suggesting that certain cognitive biases must be interpreted, from an evolutionary perspective, not as an obstacle or a shortcoming, but, on the contrary, as a

Print ISSN: 1390-3861 / Electronic ISSN: 1390-8626, pp. 197-215.

cognitive advantage, or a positive aspect of reason ("It's not a bug; it's a feature!"). This shift with respect to the Kahneman and Tversky paradigm is decisive, from the point of view of the internal consistency of the theory, insofar as it answers the question that was mentioned that operates as a trigger for the reflections that led to formulating the TAR, namely: How is it possible that an instrument so limited in its operations and crossed by biases has been adaptive and, consequently, naturally selected? The response of Mercier and Sperber (2017), in relation to this question, is simple: the adaptive character of human reason lies precisely in (some of) those characteristics that the Bounded Rationality paradigm considers as shortcomings, but that, when they are restored to the correct evolutionary landscape, they are shown as positive and beneficial characteristics for the original function for which the reason was selected, namely: to argue to persuade (and not to seek truth or knowledge).

This shift, however, is interesting for another reason, this time of a pragmatic nature: insofar as it operates as a corrective to the epistemic pessimism typical of the Bounded Rationality paradigm, it allows to give meaning again to the design of pedagogical strategies tending to improve the epistemic practices and habits of the subject —something that was virtually meaningless if one started from the overwhelming pessimism of that paradigm. This does not imply, of course, a return to the optimism typical of the Virtue Epistemology, which starts from a quasi-Rousseauian Illustrated conception of the subject, a subject that would not be traversed by cognitive, moral, or political biases, and would be guided, at least most of the time, out of the desire for knowledge and truth. None of this excessive, naive, and, to a certain extent, willful trust in man's rational capacities will be restored by TAR. What does open up is the challenge of thinking about pedagogical strategies that start from the fact that there is a certain cognitive structure (crossed by biases and unconsciously guided, most of the time, by the need to persuade) that is the result of an evolutionary process, that can make use of precisely those characteristics, instead of ignoring them, and put them to work in scenarios that, among other things, replicate the characteristics of the evolutionary landscape of human reason.

At this point, two perspectives are fundamentally opened (divergent but complementary to each other): one constructive and the other destructive. The constructive perspective outlines, to a certain extent, the conditions under which a certain didactic strategy can be effective in relation to the objective of stimulating the argumentative capacities of the participants —or, at least, it forces us to reflect on those conditions,

Print ISSN: 1390-3861 / Electronic ISSN: 1390-8626, 8626, pp. 197-215.

instead of merely supposing that any didactic strategy is effective by the mere fact of appealing to dialogue, self-criticism or the search for reasons.

The first corollary that derives from the premises proposed by TAR is the fact that when arguing, the quality and solidity of the arguments will depend on the audience one is faced with, on how counter-argumentative the interlocutor is: to produce good arguments, it takes the presence of a critical interlocutor, who pushes to produce solid, convincing arguments. The mere act of defending a certain position in an exposition before a group, for example, becomes completely unproductive if the subject knows in advance that the position presented is not going to be openly objected and questioned. The depth and solidity of the arguments, in short, will be in direct function with the critical and active interaction with the interlocutors[8].

The second corollary, and here is what is fundamental from the constructive point of view, is that the only way to overcome or counteract the negative effect of the cognitive biases of human reasoning is by putting them to work in our favor, which can be achieved, basically, in two ways: stimulating proactive reasoning and designing spaces for confrontational argumentative debate. What is meant by proactive thinking? According to Mercier and Sperber (2017), when individuals reason in isolation they tend at times, and depending on the scenario they know they will face when exposing their convictions, to emulate an agonistic argumentative context, anticipating a possible dialogical context and trying to find arguments that confirm their own position. Proactive thinking is precisely that exercise of reasoning, in an isolated and individual way, looking for arguments in favor of one's own beliefs, as if one were arguing with others, and anticipating their objections and thinking in advance of their answers. It is clear, of course, that this is not what is done all the time, but only when it is known or suspected that one is going to be faced with a situation in which they will be required to realize their own reasons. The second strategy suggested by the authors to put to work the cognitive biases in favor of oneself consists of the construction of spaces for argumentative debate —spaces that tend, by their own dynamics, and provided that they are organized by a competent mediator, to transform the confirmation bias into a tool for the logical display of the reasons that support the position of each of the participants.

But TAR contributes to understanding not only what kinds of scenarios and strategies are really effective in stimulating argumentative rationality, but also which ones are not —and that is where what can be called the negative perspective comes into play. In the first place, if the

Print ISSN: 1390-3861 / Electronic ISSN: 1390-8626, pp. 197-215.

confirmation bias is an effective characteristic of human reason, (and if, additionally, there are reasons that allow it to be interpreted as a positive characteristic of reason —at least within certain contexts), then any didactic strategy constructed on the idea of a self-critical thought seems to be destined to fail in view of the cognitive architecture that has been inherited from the ancestors: few subjects, according to these premises, spontaneously question their own beliefs. In general, for convictions to be reviewed, it is necessary that one of the following scenarios occurs: that they conflict with the beliefs of another subject, or with some characteristic of the scenario in which they are immersed, either because someone forces one to account for them, or because they collide in some way with reality, or, finally, because they are no longer effective, that is, because they no longer produce the effects that they produced until now[9]. Second, from the fact that human reasoning is particularly effective in argumentative contexts, it does not follow that *any* group discussion strategy can be effective in itself. As Mercier and Sperber (2017) state:

> When participants have clearly aligned convictions from the start, this leads to polarization. When subjects begin the discussion with ideas that are in conflict with each other and do not have a shared goal, this tends to exacerbate the differences. Group discussion is typically beneficial when participants have different ideas and a shared goal (p. 334).

Taken together, the positive indications and the aforementioned restrictions put on the stage the need to attend to the specificity of the artificial argumentation scenarios designed for pedagogical purposes, and the need to understand the spontaneous dynamics of the argumentative capacities themselves. Human reason, after all, is not a general and universal resolution module, but a specialized module (or a set of them) that arose within a specific evolutionary scenario. Failure to take these characteristics into account when designing didactic strategies can only lead to the design of naive and ineffective pedagogical strategies and, at the end of the day, to the failure of the dreams of the Enlightenment to which virtue epistemologists continue to cling.

Conclusions

Attention should be paid, as a final consideration, regarding a last point that is not at all exclusive to TAR but concerns the didactics of philosophy globally considered, and is that of the complementation between the theoretical framework and empirical support: TAR offers not only a

Print ISSN: 1390-3861 / Electronic ISSN: 1390-8626, 8626, pp. 197-215.

general theoretical framework (that of rationality as a product of natural selection) but also a host of studies from the field of experimental psychology that corroborate, at least provisionally, the predictions of the theory. The design of pedagogical strategies based solely on theoretical assumptions about human rationality, but without any type of empirical research that supports its possible effectiveness, seems doomed to walk the path followed, by way of example, by a didactic strategy such as brainstorming, a strategy whose marked ineffectiveness in stimulating argumentative debate and in leading to the search for new solutions seems to have already been clearly demonstrated[10]. Dismissing the contributions of other disciplines in the design of didactic strategies, in this sense, no longer seems to be a recommendable modus operandi, and this becomes particularly decisive in relation to the contributions of psychology and, fundamentally, of experimental psychology: What is the best way to ensure a participant's commitment to the debate after their position has been openly questioned? What kinds of attitudes do subjects tend to adopt when faced with aggressive debate scenarios? What are the effective benefits in this regard of ensuring a respectful and tolerant space for debate? What are the ways in which subjects usually resolve cases of cognitive dissonance in group discussion scenarios, where a quick response is required?

This type of questions are essential in designing effective strategies, and can only be satisfactorily answered by an approach that offers, first of all, a solid theoretical framework and articulated with the rest of the disciplines (humanistic and non-humanistic) in terms of the concept of human rationality, and, second, considerable, renewed and dynamic empirical support. TAR meets both requirements. It is not the only alternative available, of course. But it seems to be a solid, plausible, and highly flexible platform to rethink our teaching practices and the teaching strategies that we institutionally implement.

Notes

1 Although Sosa 1980 is usually considered as the touchstone of this first aspect of the Virtue Epistemology, Sosa (2011) and Greco (2010) represent two more systematic and accessible entry routes to its general guidelines.

2 Within this second aspect, Zagzebski (1996) represents, to a large extent, the founding text, both from a methodological and thematic point of view. Roberts and Wood (2007) and Baehr (2011) constitute the two most recent reference systematic approaches, in addition to the compilation by Fairweather and Zagzebski (2001).

3 Roberts and Wood (2007) offer a synthetic but comprehensive and programmatic definition of this second aspect of the Virtue Epistemology: "Virtue epistemolo-

Print ISSN: 1390-3861 / Electronic ISSN: 1390-8626, pp. 197-215.

gy, as we understand it, explores dispositional properties of persons that bear on the acquisition, maintenance, transmission, or application of knowledge and allied epistemic goods such as truth, justification, warrant, coherence and interpretative fineness. Personal traits that regularly promote such goods are virtues, and ones that impede or undermine them are vices. Relevant dispositional properties are of at least two kinds. […] It is an a posteriori normative conceptual discipline; it aims to describe knowers at their best, so it describes an ideal" (p. 257).

4 Cf. in this regard Battaly (2014).

5 Roberts and Wood (2007) represents a paradigmatic example of this emphasis.

6 The central core of the TAR is developed in Mercier and Sperber (2017 and 2019). Although there is a decisive distance between both texts with respect to the dual models of explanation of human action and a shift towards models such as social intuitionism (such as that defended in Haidt, 2001), there are no essential differences regarding the interpretation that the authors propose the axes discussed in these pages. Additional projections of the theory are found in Mercier (2011 and 2019) and Mercier and Heintz (2014). The translations by Mercier and Sperber (2017) are, in all cases, by the author of this article.

7 It is not by chance, in this sense, that Mercier and Sperber (2017) explicitly link, in the most recent exposition of the TAR, the concept of reason with that of 'reputation': "The reputation of a person is, to a large extent, the continuous effect of a conversation unfolding in time and social space about the reasons of that person. By giving our reasons, we aspire to participate in the conversation about ourselves and defend our reputation. [...] Giving reasons to justify one's actions and reacting to the reasons offered by others is, first and foremost, a way of establishing reputations and coordinating expectations" (pp. 142-143).

8 What is at stake in this statement is nothing other than the epistemic need to logically navigate each path to its ultimate consequences, which inevitably refers to the Popperian idea of letting our hypotheses die instead of us, and represents, of somehow, a mirror image of the most important premise of Socratic methodology, namely, forcing the interlocutor to unfold the idea to the maximum, until finding its limit, and, eventually, its internal contradiction or its conflict with other ideas defended by the subject. As both Socrates, Popper, and Mercier and Sperber understood, this is something that can only happen within the dialectic proper to the argumentative conflict (either real or through the dynamics of the scientific-philosophical enterprise). Finally, as Mercier and Sperber (2017) point out, this dispels the myth of the 'solitary genius': the great achievements of reason have never been the product of an individual mind, but rather a collective product, the result of interaction, always conflictive, between various individuals over many generations (pp. 315-327).

9 This explains, incidentally, the exceptional character of the capacity for self-criticism in highly hierarchical relationships, on the part of those who are in the position of power. As Mercier and Sperber (2019) point out, "many of our beliefs are prone to remain unchallenged because they are only relevant to ourselves and we do not share them, or because they are controversial only with the people we interact with, or because we have sufficient authority to affirm them" (p. 26).

10 Cf., by way of example, Diehl and Stroebe (1987) and Mullen, Johnson and Salas (1991).

Print ISSN: 1390-3861 / Electronic ISSN: 1390-8626, 8626, pp. 197-215.

Bibliography

BAEHR, Jonathan
2011 *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology*. Oxford: Oxford University Press.

BARGH, John & CHARTRAND, Tanya
1999 The unbearable automaticity of being. *American Psychologist*, *54*(7), 462-479. https://doi.org/10.1037/0003-066X.54.7.462

BATTALY, H.
2014 Acquiring Epistemic Virtue: Emotions, Situations, and Education. En A. Fairweather, O. Flanagan (Eds.), *Naturalizing Epistemic Virtue* (pp. 175-196). Cambridge: Cambridge University Press.

DIEHL, Michael & STROEBE, Wolfgang
1987 Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, *53*(3), 497-509. https://doi.org/10.1037/0022-3514.53.3.497

EVANS, Jonathan
2008 Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255-278. https://doi.org/10.1146/annurev.psych.59.103006.093629

EVANS, Jonathan & STANOVICH, Keith
2013 Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, *8*(3), 223-241. https://doi.org/10.1177/1745691612460685

FAIRWEATHER, Abrol & ZAGZEBSKI, Linda (Eds.)
2001 *Virtue Epistemology: Essays on Epistemic Virtue and Responsibility*. Oxford: Oxford University Press.

GIGERENZER, G.
2008 Moral intuition = Fast and frugal heuristics? En W. Sinnott-Armstrong (Ed.), *Moral Psychology. 2. The Cognitive Science of Morality: Intuition and Diversity* (pp. 1-27). MIT Press.

GRECO, John
2010 *Achieving knowledge: A virtue-theoretic account of epistemic normativity*. Cambridge: Cambridge University Press.

HAIDT, Jonathan
2001 The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814-834. https://doi.org/10.1037//0033-295X.108.4.814

KAHNEMAN, Daniel
2012 *Pensar rápido, pensar despacio*. Madrid: Debate.

KAHNEMAN, Daniel & TVERSKY, Amos
1979 Prospect Theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263-291. https://doi.org/10.2307/1914185

MERCIER, Hugo
2011 What good is moral reasoning? *Mind & Society*, *10*(2), 131-148. https://doi.org/10.1007/s11299-011-0085-6

Sophia 30: 2021.

Print ISSN: 1390-3861 / Electronic ISSN: 1390-8626, pp. 197-215.

MERCIER, Hugo & HEINTZ, Christophe
2014 Scientists' Argumentative Reasoning. *Topoi*, *33*(2), 513-524. https://doi.org/10.1007/s11245-013-9217-4
MERCIER, Hugo & SPERBER, Dan
2017 *The Enigma of Reason: A New Theory of Human Understanding*. Cambridge: Harvard University Press.
2019 ¿Por qué razonan los humanos? Argumentos para una Teoría Argumentativa (C. McDonnell & J. M. Vivas, Trads.). *Cuadernos Filosóficos / Segunda Época*, *15*. https://doi.org/10.35305/cf2.vi15.54
MULLEN, Brian, JOHNSON, Craig & SALAS, Eduardo
1991 Productivity loss in brainstorming groups: A meta-analytic integration. *Basic and Applied Social Psychology*, *12*(1), 3-23. https://doi.org/10.1207/s15324834basp1201_1
NICKERSON, Raymond
1998 Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175-220. https://doi.org/10.1037/1089-2680.2.2.175
ROBERTS, Robert & WOOD, Jay
2007 *Intellectual virtues: An essay in regulative epistemology*. Oxford: Oxford University Press.
SIMON, Herbert
1955 A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, *69*(1), 99-118. https://doi.org/10.2307/1884852
SOSA, Ernest
1980 The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge. *Midwest Studies In Philosophy*, *5*(1), 3-26. https://doi.org/10.1111/j.1475-4975.1980.tb00394.x
2011 *Knowing Full Well*. Princeton: Princeton University Press.
STANOVICH, Keith & WEST, Richard
2000 Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), 645-665. https://doi.org/10.1017/S0140525X00003435
STANOVICH, Keith & WEST, Richard
2008 On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*(4), 672-695. https://doi.org/10.1037/0022-3514.94.4.672
TVERSKY, Amos & KAHNEMAN, Daniel
1981 The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453-458.
ZAGZEBSKI, Linda
1996 *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge: Cambridge University Press.

Document reception date: July 15, 2020
Document review date: September 15, 2020
Document approval date: October 15, 2020
Document publication date: January 15, 2021

Sophia 30: 2021.

Print ISSN: 1390-3861 / Electronic ISSN: 1390-8626, 8626, pp. 197-215.