INTENTIONAL ATTRIBUTION AND RATIONALITY: A CRITICAL READING OF DENNETT'S INTENTIONAL ATTRIBUTION PROGRAM

EDGAR ESLAVA
Profesor Facultad de Filosofía y Letras, Universidad Santo Tomás, Bogotá-Colombia

INTENTIONAL ATTRIBUTION AND RATIONALITY: A CRITICAL READING OF DENNETT'S INTENTIONAL ATTRIBUTION PROGRAM

Abstract: In this paper I introduce some of the key elements of Daniel Dennett's theory of intentional attribution and their relation with his notion of rationality. While doing so I will show that Dennett's approach implies a circularity in the process of attribution of rationality, and that his resource to evolutionary arguments for trying to avoid an infinite regress does not help him to avoid the problem. My presentation will include a revision of Dennett's arguments for epistemic intentional ascription and rationality attribution as well as some criticisms developed against his proposal. At the end of the article I will extend the criticisms and present my view about his proposal for ideal rationality attribution.

Keywords: Rationality, intentionality, Dennett, consciousness, intentional attribution.

ATRIBUCIÓN DE INTENCIONALIDAD Y RACIONALIDAD: UNA LECTURA CRÍTICA AL PROGRAMA DE ATRIBUCIÓN INTENCIONAL DE DENNETT*

Resumen: en este texto se presentan algunos de los elementos centrales de la teoría de la atribución intencional de Daniel Dennett y sus relaciones con la noción de racionalidad. Se mostrará que la aproximación propuesta lleva implícita una circularidad en el proceso de atribución de intencionalidad, y que su uso de argumentos evolucionistas para intentar evitar un regreso al infinito no le son suficientes para evitar el problema. La presentación concluye con una revisión de los argumentos usados por Dennett para la adscripción de intencionalidad epistémica y atribución de racionalidad, así como algunas críticas alrededor de esta propuesta. Al final del texto se adelantan algunas críticas propias frente a la noción de atribución de racionalidad ideal propuesta del autor norteamericano.

Palabras clave: racionalidad, intencionalidad, Dennett, conciencia, atribución de intencionalidad.

Fecha de recepción: Julio 7 de 2015
Fecha de aceptación: Mayo 18 de 2016

Forma de citar (APA): Eslava, E. (2016). Intentional attribution and rationality: critical reading of Dennett's intentional attribution program. *Revista Filosofía UIS*, 15 (1), 225-243, doi: http://dx.doi.org/10.18273/revfil.v15n1-2016011

Forma de citar (Harvard): Eslava, E. (2016). Intentional attribution and rationality: critical reading of Dennett's intentional attribution program. Revista Filosofía UIS, 15 (1), 225-243.

Edgar Eslava: colombiano. Profesor Facultad de Filosofía y Letras, Universidad Santo Tomás, Bogotá. Licenciado en Física Universidad Pedagógica Nacional, Magister en Filosofía, Pontificia Universidad Javeriana, Ph.D. en Filosofía, Southern Illinois University.

Correo electrónico: edgareslava@usantotomas.edu.co.

* Revisión de tema

INTENTIONAL ATTRIBUTION AND RATIONALITY: A CRITICAL READING OF DENNETT'S INTENTIONAL ATTRIBUTION PROGRAM[1]

1. Presentation

The main problems a theory of mind should address, one can say paraphrasing Dennett, are the *what* and the *how* of Consciousness and Content (1978, pp. 351-366). Only by answering those questions would a theory of the mind possibly aspire to be considered successful. Dennett's proposal for such a theory that responds properly to the questions of what contents can be rightly assigned to an entity and how such contents are arranged and interconnected is developed in his Intentional Attribution program. Whatever minds could be, and whatever our mental states are, Dennett would say, the only way we can be aware of them is by the study of the way such things play their roles in the interactions between subjects and their environments, i.e. their behavior. Dennett's Intentional Attribution program is a behavioral predictive strategy that allows us to anticipate the actions of the subjects based on the information we have about the structure that underlies their behavior. Dennett's thesis is that intentional attitudes such as beliefs, desires, pains, and so are "perfectly objective phenomena", whose existence can only be discerned from the point of view of a researcher that adopts certain predictive strategy, the intentional attribution strategy (1979, pp.13-36).

Dennett's three-stance model

In order to show how the prediction of behaviors is not only possible but also effective, Dennett introduces his so-called three-stance model, that operates this way: in the most basic level of reality, things (creatures, entities) are simply material aggregates that differ from one another only in their degree of complexity.

[1] This paper is a result of the research project Epistemic and non-epistemic beliefs, led by the author.

When dealing with entities at this level, it is enough to understand the physical laws that rule the interaction among its parts to be able to predict the entity's behavior. In such a case one's perspective is what Dennett named the Physical Stance, a perspective that allows us to predict entity's behavior limited only by our knowledge of the building blocks and basic laws of the material world. The Physical Stance is a perspective applicable to any entity; but only in the case of the very simple ones will its use lead us to reliable results. It is perfectly possible to anticipate the trajectory, physical behavior, of a rock or an arrow by using only the Newtonian laws of motion, and the results are going to be as accurate as our knowledge of the laws permits. But not all entities are as simple as rocks and arrows. In the case of ballistic missiles for example, their behavior can be anticipated not only by assuming they are physical compounds, but also by the assignation of some structural design. In this case, the prediction of behavior would rest on the reliability of the design and the optimal physical structure of the entity. For this reason, Dennett called this stage the Design Stance. In this level, prediction relies on the combination of the physical and design features, and in the idea that the entity would act in a way that would not prevent it from fulfilling the task it was designed for.

But the complexity of the world does not stop at the design level; some complex machines and every living creature (or at least most of them) are much more complex in their behaviors than pre-established trajectory followers. Dennett's proposal is that, in order to make sense of the behavior of such complex systems, it will be useful to treat them as if they were rational agents whose decisions are based on beliefs and desires, in other words, as intentional agents. When trying to predict the behavior of such systems, one has to assume them operating as optimally designed devices, device that would do their best to achieve any goal they are pursuing. By using this strategy, one is approaching behavior's prediction from the Intentional Stance.

Intentional attribution is a bi-directional strategy that operates either from simplicity to complexity or from complexity to simplicity. On the ascendant side one starts at the physical stance in order to predict entity's behavior, if one generates a successful predictive model based on the mere physical structure of the subject, then one stops right there. But if there are some conducts of the entity that are not possible to be anticipated from this perspective, then it would be necessary move into the design stance and to treat the subject as a physically optimal structure whose design allows it to behave in a particular way. Again, if this is enough for obtaining a reliable prediction no more steps are needed, otherwise, one final move is required. One starts treating the entity as a rational agent whose responses are the result of presenting and contrasting alternative courses of action and choosing the most effective among them. Even the highly complicated behavior of humans can be predicted reliably by using this strategy.

On the other side of the system, when facing complex subject's behavior one starts by ascribing entities with intentional attitudes and use them as predictive tools: "one stars with the ideal perfect rationality and revises downward as circumstances dictate" (Dennett's, 1979, pp. 21) What is particular in this case is that one is required to descend to the design stance only if the attribution of rationality is not guaranteed. If one is going to attribute intentions to a table or a chair, one is going to end ascribing it things like its desire to serve as support, or its believing to be the center of the world. This would offer no help when trying to anticipate if it would hold a very heavy refrigerator during a Saturday morning cleaning session. The same logic rules the passage from the design stance to the physical stance, asking for the "design" features of an electron is going to tell us nothing about, say, its electric charge. In cases like those, intentional ascription would introduce more problems than solutions, making complexity out of simplicity.

It has to be noticed that Dennett's intentional attribution system depends on the assignation of rationality to the subject. The subject has to believe all the consequences of its actions in order for its behavior to be predictable. This commitment is so strong that if the predictive system fails, the causes of the failure have not to be tracked down to some irrationality in the entity's behavior, but to the predictive system itself.

> The presumption of rationality is so strongly entrenched in our inference habits that when predictions prove false, we at first cast about for adjustments in the information possession conditions or goal weightings, before questioning the rationality of the system as a whole (Dennett 1978, p.9).

Making explicit Dennett's notion of rationality and its implications to the intentional attribution program are tasks I am going to deal with latter in this text.

2. Intentional attribution and rationality

Despite the presumption of rationality, beliefs themselves, along with the other intentional attitudes, are attributed to entities just as a pragmatic tool, not as any sort of metaphysical notion. Such assumptions of rationality are nothing more than "heuristic idealizations, justifiable only insofar as they lead, by and large, to true intentional stance predictions" (Fodor, 1981, p. 105). Dennett has defended this project in all his papers about intentional attribution. As an example, let's see it as part of a discussion about one of Dennett's favorite cases:

> Lingering doubts about whether the chess-playing computers *really* have beliefs and desires are misplaced; for the definition of intentional systems I have given does not say that intentional systems *really* have beliefs and desires, but that one can explain and predict their behavior by ascribing

> beliefs and desires to them, and whether one calls what one ascribes to the computer beliefs or belief-analogues or information complexes or intentional whatnots makes no difference to the nature of the calculation one makes on the basis of the ascriptions (1971, p. 7).

The attribution of intentional attitudes to an entity depends on the idea that what count as beliefs, desires, and the so, are "all the truths relevant to the system's interests (or desires) that the system's experience to date has made available" (Dennett, 1979, p.19). This makes true believers, the only entities whose behavior is predictable from the intentional stance, believers of truths.[2]

Being strongly committed with the mentioned rationality requirement for the subject, Dennett's is interested in showing a reliable source for rationality, one that sets the intentional attribution program in solid ground. Natural Selection offers a promising answer, and Dennett is eager to use it. According to Dennett's view on natural selection, living creatures are organized in a continuous chain that goes from simplicity to complexity. Such a chain, an "evolutionary chain" in the Darwinian sense of the word, starts with the basic organisms whose response to the environment is determined by natural selection. These so-called Darwinian Creatures (1996, p. 83) develop responses in a blind way, just as the result of an endless procedure that generates 'mutations' in the offspring of the current generation. In the long run, mutations will allow the survival of the species by making apt individuals, those with the "right" features", to survive and be able to reproduce. The next step in the chain is the one of the creatures whose responses are not blind or forced by the laws of natural evolution, but advanced out of a stimulus-response relation with the environment. Skinnerian Creatures learn what

[2] I would like to make a brief comment about the parallel between Dennett's ascription of truth to the beliefs and Davidson's principle of charity for radical interpretation. Trying to solve the question about the ability language interpreters exhibit when making sense of someone else's utterances, Davidson proposes to "take the fact that the speakers of a language hold a sentence to be true (under observed circumstances) as prima-facie evidence that the sentence is true under those circumstances" (Davidson, 1974). This principle of charity, or Charitable Principle, is the very central axiom of Davidson's system and encloses important elements of the interpretation attempt. It makes the truth of an utterance relative to "observed circumstances". Those circumstances are in turn dependent on the moment and context of the performance of the utterance. Therefore, the truth or falsity value of a sentence is something to be judged only by somebody who knows the actual situation under which the utterance was done. But because the interpretation is a two-way process, it also implies that the speaker must know that the conditions under which his utterances are made are such that the listener has a good chance to make sense of them. In other words, the principle of charity defines the conditions under which the linguistic behavior of the agents is necessarily rational. In Dennett's case, those conditions are also going to be the ones that would allow interpreters to make sense of the behavior of concrete entities' behavior, under appropriate circumstances, i.e., under the basic assumption of subject's rationality.

kinds of responses are the adequate ones to the exigencies of the environment they happen to be immersed in, and act according to such exigencies. These entities are capable of developing behaviors based on the specific information provided by the particular environment they live in, but are still unable to make any kind of prediction about the future course of actions in the surrounding world. Such an interesting feature is what characterizes beings belonging to the chain's third step, where the anticipation of future events is what generates alternative sets of actions. Those alternative sets are developed by contrasting possible courses of action in mental environments of representations of the world, in which favored behaviors are contrasted against mental representation instead of against the world itself, in a process that allows the more imaginative individuals to get a better chance to succeed. Such a capacity for imagining and comparing is no doubt a remarkable one, but if one thinks this is an amazing system for the generation of behaviors, there is still another much more impressive system: "meta-learning." At the top of the "evolutionary chain" there are creatures capable not just of reacting to the forces coming from the environment, or of learning how and when to generate specific kinds of responses but that are even able to learn about their own ability for learning and improve it. These creatures have developed the competence for manipulating the incoming information for the creation of "smart moves," behaviors designed to maximize the probability of success. Of course we human beings occupied this place. Is in this sense that Dennett sees us as *optimal creatures*, creatures whose abilities integrate and exceed those of the inhabitants of the other levels, and whose responses are based on well studied and elaborated analysis of the environment and of the subject's performances capabilities.

However the features of the creatures in the top of the layer may appear particular to humans, for Dennett it is clear that we are not the only ones inhabiting the place. There is a lot of evidence that beings other than humans have the ability to develop strategies for "smart moves" out of the information they receive from the environment. As mentioned before, Dennett's favorite examples are chess-playing computers; such devices have as part of their programs the skills for choosing from a set of possible moves those that are better responses to the actual situation of the game than others from the set of possible ones. Such a capability, either present since the primal programming of the machine or developed out of the (also designed) faculty of learning from past moves or games, matches perfectly the criteria presented for being considered an optimal creature, making the "chain of living creatures" extend to the non-living ones.

The move toward natural selection has paid then the expected dividends and more. It has made intentional attribution program, as a tool for making sense of creature's behavior, continuous with natural evolution theory as an epistemic approach to the behavior (organization and speciation) of living organisms. It has also permitted the connection between living and non-living creatures as subjects whose behavior can be understood and anticipated from the intentional stance.

Despite the results Dennett's intentional attribution system could provide us with, there are some elements of the intentional attribution program that would require further analysis. Among them, there is the relation between researcher's intentional states and subject's intentional states, a relation that seems more complicated than what Dennett seems to admit. Dennett requires agents to be perfectly rational in order its behavior to be predictable. But what counts as a guarantee for the attribution of rationality? The rest of this paper will be devoted to the task of sharpening this question and the searching for some answers. To do so, I will present first my arguments against Dennett's view of the relation between rationality and natural selection. Then I will challenge the notion of ideal rationality of the agents that underlies the Dennett's general program. It has to be noticed right now that by no way such criticisms are brand new; a lot has been written about those elements of the intentional attribution program. For such a reason my approach will follow prior critical views, both because there are some shared elements between them and my personal view and because of the lightening perspective they offer. However, at least I expect, my own criticism would take distance from the prior ones, showing new problems and presenting new challenges to Dennett's program.

3. Natural selection and intentionality

Let's start with Dennett's approach to natural selection. Dennett's theory of mental content can be described as a functional approach, one in which

> All attributions of content are founded on an appreciation of the *functional roles* of the items in question in the biological economy of the organism (or the engineering economy of the robot). This is a specifically "teleological" notion of function (not the notion of a mathematical function or of a mere "causal role" as suggested by David Lewis and others). It is the concept of function that is ubiquitous in engineering, in the design of artifacts, but also in biology (1998, pp.359-360).

From such a functionalist perspective, Dennett's general program deals with subject's reasons *via* idealization of instrumental rationality. But in his approach to natural selection and the way subjects become rational, Dennett uses the passage from simplicity to complexity as a teleological guide. And such a notion of teleological rationality permeates Dennett's entire program, compromising it. In other words, Dennett needs natural selection to warrant the necessity of perfect rationality. This is what Stephen Stich has named Dennett's argument from natural selection.[3] "But why," asks Stich, "would the mere existence of natural selection suffice to ensure that the creatures would be good approximations to the thoroughly rational ideal embodied in the notion of intentional system? (1984, p.256) It is just not true that natural selection operates favoring inferential strategies that yield

[3] (Stich, 1984).

mostly true beliefs, and Dennett seems to forget it. What natural selection favors is the appearance of beliefs that, in the long run, would generate some selective advantage.

> Natural selection may often favor a process that yields false beliefs all the time, but which has a high probability of yielding true beliefs when it counts. […] If eating a certain food causes illness on a single occasion, the organism would immediately come to believe (falsely, let us assume) that all passingly similar foods are poisonous as well. When it comes to food poisoning, better safe than sorry is a policy that recommends itself to natural selection (1984, p. 126).

What Stich is pointing out is the fact that is it not necessary that evolution produces only successes. A mutation, a blind change, is not *per se* a mistake or a hit; it is only a change. In such a sense, single mutation could be as right as it could be mistaken, but that is something to be evaluated *a posteriori*, after a large number of replications of the new features. And this will happen far in the future of the original change. This is precisely what happens all the time, only some changes increase organisms' opportunities for survival, but even the most extravagant mutation could generate evolutionary gain. If there are some dramatic changes in the environment it is possible only the organisms which such a dramatically new feature would survive, making an odd change the best of the options. In this respect, Dennett operates with a very rigid notion of change, and nature has shown to be more dynamic than that.

Dennett answers, claiming for the necessity of the evolutionary argument for any interpreter engaged in the intentional attribution program:

> We have already seen that there is no point in ascribing beliefs to a system unless the beliefs ascribed are in general appropriate to the environment, and the system responds appropriately to the beliefs. An eccentric expression of this would be: the capacity to believe would have no survival value unless it were a capacity to believe truths. What is eccentric and potentially misleading about this is that it hints at the picture of a species "trying on" a faculty giving rise to beliefs most of which are false, having its inutility demonstrated, and abandoning it. A species might "experiment" by mutation any number of inefficacious systems, but none of these systems would deserve to be called belief systems precisely because of their defects, their non-rationality, and hence a false belief system is a conceptual impossibility (1978, p.17).

The answer is then that we have to stand on the solid ground of our own interpretation system in order to make sense of the role played by evolution in the process of selecting true believers. But even if this point is conceded in favor to Dennett, the question about Natural Selection's role as the designer of living organisms would still be open. Beyond the role of our own interpretative system nothing has been said yet about the way natural selection accounts for the

presence of rationality as a common feature to some of the organisms in the final steps of the evolutionary chain.

In an attempt to clarify his position about this issue, Dennett draws an analogy between his account of Natural Selection as a designer and R. Dawkins's model of human organisms as survival engines for genes[4]. Let us imagine, Dennett says, a device designed to hold a receptacle containing the body of a person who wants to hold his life in suspension for four centuries and to be "waked up" at the end of that time lapse. Such a device would have to be designed to provide the necessary energy and protection to the receptacle in order to fulfill its final task, i.e., the preservation of the live that lies beneath its mechanical armor. Among the multiple designs one can ask for, the most effective one would be that capable of recognizing the environment and respond to that knowledge acting according to its main purpose. Mobility in order to avoid threatening intruders and to pursue energy supply would be one of the expected features for the engine to posses will it increase the survival chances of its precious load. A big robot in whose interior the receptacle has been placed safety would be a vivid image of the desired system. But not any robot would be able to accomplish the task. As one of the multiple individuals in a highly populated environment, the robot must be prepared to compete with other individuals for the basic supplies required to warrant its proper functioning, as an indispensable requirement to complete its task.

> It would no doubt be wise to design it [the robot] with enough sophistication in its control system to permit it to calculate the benefits and risks of cooperating with other robots, or forming alliances for mutual benefits. The result of this design project would be a robot capable of exhibiting self-control … capable of deriving its own subsidiary goals from its assessment of its current state and the import of that state for its ultimate goal (which is to preserve you) (Dennett, 1987, p. 297).

The kind of robot just described would be the perfect engine for the assigned job. But if it seems that such an engine is just the artificial intelligence engineers' dream, Dennett reminds us that this is just a variation on the theme of biological species as hosts, survival engines, of some astute and selfish entities, genes. Dennett's account of Dawkins model states that living creatures, with human beings among them, have been designed as survival engines for genes. The very limited capability of such entities for interacting with the environment would make its subsistence simply impossible, unless they would count with the unmeant cooperation of the creatures that host them, us, and whose interests we pursue while pursuing ours.

However appealing this analogy would seem, Dennett acknowledges quickly, it is nevertheless incomplete:

[4] Dawkins (1976) and Dawkins (1986)

> In my tale I supposed that there was conscious, deliberated, foresighted engineering involved in the creation of the robot, whereas even if we are, as Dawkins says, the product of design process that has our genes as the primary beneficiary, that is a design process that utterly lacks a conscious, deliberate, foresighted engineering (1987, p. 299).

The discrepancy between the two cases is obvious, it rests in the question of who is responsible for the design of biological engines. Dennett's own answer to this question is that Mother Nature, "the long slow process of evolution by Natural Selection" has played such an outstanding role. The main bulk of Dennett's article is devoted to the justification of his proposal of natural selection as designer of living creatures, but I am not dealing with such a justification in this paper. My purpose is just to show that there is a second limit in the analogy that Dennett seems to have left untouched.

The difficulty I want to point out arises from the tension between Natural Selection's blindness, a fact Dennett stresses widely in his philosophical work, and the necessity for a teleologically oriented design as part of the model just presented. In his example of the robot and the receptacle the main task of the machine is defined since the moment of its conception, namely, to preserve the life of the individual inhabiting the receptacle. All the secondary goals the robot is able to define in order to guarantee energy supply and security are just subsidiary to its main goal, something clearly defined since the creation of the robot. Without the basic aim of preserving the live inside the capsule, something decided by the designer, the robot's secondary goals are superfluous, as it is the engine itself. It would make no more sense for our robot trying to search for energy sources than for a ballistic missile to fly without a defined target. And here is where the tension arises, because for the analogy to be held whatever the designer ends up being, it would have to be working under the idea of designing toward a specific end. Without a clear end there would not be any mandatory task to complete, and then no reason for the engine to exist. Of course this is not the case for Natural Selection. Acting over changes produced randomly and not by any prior motivation, Natural Selection does not pursue any specific goal but "chooses" (to use one of Dennett's analogies) among the changes those that increase species' chances to reproduce effectively and survive. In other words, Natural Selection with "original reasons" would no longer be blind. And blindness is for Natural Selection its most precious quality. Then, the analogy does not seem to run as expected and the conclusions Dennett obtains from its use would have to be re-evaluated, or the system that rest on Natural Selection as an explanatory tool would be jeopardize.

However, in order to be fair to Dennett one has to admit that in every new generation of a species, the goal of Natural Selection is to ensure that the creatures best suited to the environment are precisely those whose chances to reproduce successfully are bigger. Such a goal, one would be tempted to say, is specific

enough to keep the analogy working, and with it the desired conclusions.[5] In defense of my case, I will point out that the fact that some creatures survive due to their particular features, those generated blindly by mutation or by genetic recombination, is a fortuitous result of natural selection, and not the necessary conclusion of some intended rational effort. At the end, what is under discussion here is the difference between "ascription" and "existence" of such an effort. To make myself clear, let me introduce another critical perspective of Dennett's approach. It is an attack to the sort of account for the role of Natural Selection as an interpretative perspective.

It is puzzling how Dennett thinks an appeal to the Darwinian theory —which is, after all, a causal story about the *mechanisms of speciation*— could reveal an "element of interpretation" in content ascription. Interpretativism is, *inter alia*, the view that, strictly speaking, we don't really have beliefs and desires. But, one supposes, what a creature *doesn't really have* can't help it much in its struggle for survival. It is for exactly this reason that, unlike Dennett, most people who take an evolutionary line on Intentionality are correspondingly Realist about content. *Qua* Darwinists, they suppose that there's a matter of fact about what selection history a creature has and about what mechanisms served to mediate its history of selection. So they are required to suppose also that organisms can't be selected for believing truths unless they do believe truths (Fodor and Lepore, 1993, p. 74).

Fodor and Lepore's target is the tension between the role beliefs and desires as just ascribed features of the subjects and their role as results and leaders of the evolutionary process. On the one hand, if they are nothing but heuristic tools that help the interpreter to make sense of the subject's behavior, then to treat them as ends promoted by evolution or anything of that sort. On the other, if there is anything at all that beliefs and desires has been selected for by nature, then it is nonsense to talk about them as simply hermeneutic idioms. I could not agree more with the authors of this criticism. Either a set of features exists or does not exist, even if it is possible, as certainly it is, to refer to and make use of ontologically nonexistent objects. But the fact that one can use such expressions does not allow us to conclude their existence, and of course not to introduce their existence as an explanation or as a conclusion of any empirical argument.

There is also a second tension Fodor and Lepore indicate in Dennett's intentional attribution program. In this new case the target is the use of the intentional stance as interpretative mean to understand evolutionary processes. The tension now is between two different roles assigned to evolution: the role of justifier of the attribution of intentional states to ourselves, and its role of subject over which the intentional strategy can be applied. For the first one, Dennett has stated that without and appeal to what "Mother Nature has in mind" no attribution of intentional

[5] I am in debt to Professor Pat Manfredi for pointing out this subject to me.

states to us can be sustained. For the second, any attribution of intentional states to Mother Nature rests on the assumption of optimal design we use to make sense of evolution. Then, Fodor and Lepore state, "apparently, the hermeneutic status of intentional ascriptions (to us) derives from the correspondingly hermeneutics status of ascriptions of biological functions (to mental states), which in turn derives from the hermeneutics status of intentional ascriptions (to Mother Nature)." The circularity of such an approach is just obvious. As in the prior case, I agree with this view of Dennett's attempt and, in fact, I find in it supportive for my argument against the proposed analogy between evolution (Mother Nature) and Dawkins' robots. The only way out of the problem of combining existent and non-existent features in the landscape is to recur to a circular definition of the way natural selection shapes species by appealing to its postulated final goals. The ascription of such final goals, as seen, has been introduced as an element of the interpretative system, making the intended conclusion of nature operating "as if" aiming some specific goal the result of the application of a circular argument.

As a result of the discussion up to this point, it can be said that, even if Dennett arguments have not been shown to be unsound, at least there are several open questions that Dennett's appealing to natural selection still has to solve.

4. Ideal rationality under fire

A second problem I will deal with is the situation which Stich has called the argument from the inevitable rationality of the believers. From Dennett's perspective, states Stich,

> When we attribute beliefs, desires and other states of commonsense psychology to a person- or for that matter to an animal or an artifact- we are assuming or presupposing that the person or object can be treated as an *intentional system*. An intentional system is one which is rational through and through; its beliefs are "those is ought to have, given its perceptual capacities, its epistemic needs, and its biography… [Its desires] are those it ought to have, given its biological needs and the most practicable means of satisfying them… [And its] behavior will consist of those acts that it *would be rational* for an agent with those beliefs to perform." The point is not that people *must* be rational. No such conclusion follows from Dennett's view. What does follow from Dennett's view is that people must be rational *if they can usefully be viewed as having any beliefs at all*. We have no guarantee that people will behave in a way that makes it profitable for us to assume the intentional stance toward them. But intentional descriptions and rationality come in the same package; there is no getting one without the other… If a system infers irrationally, it cannot be an intentional system; thus we cannot ascribe beliefs and desires to it. *But since inference is a belief generating process, the system does not infer at all* (1984, pp. 254-255).

What Stich shows here is that, from Dennett's perspective, the possibility for a subject's behavior to be understood from the intentional instance is a situation of all or nothing for the ascription of rationality to the subject. If no rationality is ascribed, then the intentional stance approach will be useless. But worse than that, if there is no chance for using profitably the intentional stance in a particular subject, the subject has to be considered as a non-inferential system at all. According to Stich, the problem with the model is the Dennett has mistaken the relation between our ordinary notions of belief and desire and his notion of an idealized fully rational intentional system. It happens not to be the case that the ordinary ascription of beliefs and desires presuppose fully rationality. In fact, Stich argues, "there is nothing in the least incoherent or unstable about a description, cast in intentional terms, of a person who has inconsistent beliefs" (255). Then, what is mandatory for the ascription of intentional attitudes to a subject from the intentional stance seems not to be necessary at all in any ordinary situation, making Dennett's system unnecessarily restrictive.

Dennett's defense to this attack is to clarify the scope and limits of the intentional stance interpretative approach. Once such a task is completed, there would be no place for problems of the sort presented by Stich. Dennett holds that nobody is always perfectly rational, invulnerable to fatigue or lack of memory, or to any sort of malfunction or design imperfection. Any of those situations would lead to circumstances where the interpretation and anticipation of behaviors by intentional attribution would be impossible, "in much the same way the physical damage to an artifact, telephone or automobile, may render it indescribable by the normal design terminology for that artifact" (1979, p. 28). Then, behavioral abnormalities cannot be considered challenges for the intentional ascription strategy, since they are, by definition, out of the scope of the strategy. In addition, states Dennett,

> It is at least not obvious that there are systematically irrational behavior or thinking. The cases that have been proposed are all controversial, which is just what my view predicts: no such thing as a cut-and-dried or obvious case of "familiar irrationality." This is not to say [again] that we are always rational, but that when we are not, the cases defy description in ordinary terms of belief and desire (1981, p. 87).

In such a case, no interpretation can be settled on. As a general result, once realized that the intentional attribution strategy applies only under the assumption of agent's rationality, the problem of misinterpreting or lack of interpretation in situation of irrationality is not a problem the intentional strategy has to solve. This is so both because such situations fall beyond its scope and because it has not been proved that something like systematic irrationality exists.

But there is no easy way out from the problem of attribution of rationality. In fact, Fodor has developed a direct attack on Dennett's use of the attribution of ideal rationality. Defending his own agenda, the existence of propositional (intentional) attitudes as real mental entities and not just as simple interpretative categories, Fodor has pointed some inconsistencies in Dennett's model.

> Dennett's analysis [of intentional attribution] explains the utility of the intentional idiom without assuming that there are facts that intentional ascriptions correspond to. If the analysis is right, then the least hypothesis is, surely, that there are no such facts and that appears to be the conclusion Dennett wants us to endorse. What, then, can be said in favor of the analysis? Dennett's main argument goes something like this: Since intentional stance predictions will work only insofar as we are dealing with rational systems, an assumption of rationality is implicit in every such prediction. But such assumptions are ours to make or to withhold: i.e., they are themselves heuristic idealizations, justifiable only insofar as they lead, by and large, to true intentional stance predictions. So, we can conclude the unfacticity of propositional attitude ascriptions from the unfacticity of the rationality assumptions that they presuppose (Fodor, 1981, p. 115).

What Dennett would have to do for his argument to be sound, states Fodor, is to be able to prove two related things. First, that in any attempt to predict the behavior of an entity from its intentional states, some assumption of rationality is implicit; second, that the character of such assumptions is nothing else but simply heuristic. Neither of them, holds Fodor, are accomplished by Dennett; and in each case the cause of the failure is possible to be traced. For the first one, Fodor states, Dennett's intentional system does not rest on an assumption of rationality, but on a counterfactual assumption of rationality. Instead of stating that because of its necessary rationality it is possible to predict subject's behavior, Dennett's defense of the intentional attribution program relies on the notion that such a behavior would not be possibly described unless irrationality were discarded from subject's behavior. As it has been shown, the introduction of the argument from natural selection is meant to hold Dennett's case. For the second of the uncompleted tasks, Fodor shows that the required notion of ideal rationality would only be necessary if realism about propositional attitudes is shown to be wrong. Otherwise, the mental representations of an organism, the mental processes that model its logic would offer a more accurate account of the subject's intentional behavior. But, Fodor insists, Dennett has continuously failed in showing the flaws of realism about propositional attitudes[6]. Then, Dennett has not fulfilled the requirements for its system to be solid, and the conclusions it holds would simply not follow. This short essay is not the place to discuss Fodor's procedures and results. For the present theme it will be sufficient to show Dennett's defense, and to see how it is intended to articulate with the central body of his interpretative system.

[6] See the results presented infra, 9.

Against these new charges Dennett's answer resembles the prior ones. He will show the limits and suitable extent of the proposed notion of ideal rationality, and use the results to justify his personal conclusions. As said, this is just like what was done for the defense of intentional attribution as a valid interpretative system. The first step towards the wanted clarification is Dennett's own recognition of the tension present in his theory. Even playing the assumption of rationality a crucial role in the intentional attribution program, he recognizes having systematically resisting any declaration on the nature of rationality itself. The main reason for doing so is that the intentional attribution program rests on a concept of rationality is systematically pre-theoretical. Rationality, Dennett says, cannot be identified with deductive closure: being rational is not believing all the logical consequences of every belief one have. Such a claim would be in open contradiction with the natural limits of real-time responses the brain is forced to produce. It would be impossible, infinitely time consuming, for a brain to compute, hold and remember all the information such a definition would require. Nor is rationality, continues Dennett, perfect logical consistency: the inexistence of any contradictory beliefs would not be possible, or even thinkable, to avoid given the mechanical efficiency of our brains, and our lack of permanent and perfect memory, attention, and knowledge. Finally,

> I am careful not to define rationality in terms of what evolution has given us- so I avoid outright tautology. Nevertheless, the relation I claim holds between rationality and evolution is […] that if an organism is the product of natural selection we can assume that most of its beliefs will be true, and most of its beliefs-forming strategies rational (Fodor,1981, p. 97).

What then does it mean that rationality is considered as a "pre-theoretical concept"? It means that whatever rationality is, it is necessarily a concept based on our "shared intuitions about what makes sense." What else, asks Dennett, could one rely on? (98) As an intuitive and non-formal concept, rationality is nothing more than a "general purpose term of cognitive approval," and this in turn, implies the revisable and conditional character of the belief-forming strategies and their resulting outcomes. Then, instead of a strait definition of rationality, or of rationally formed beliefs, Dennett's morals is that all we can do when judging the rational status of any agent's beliefs and desires is to make comparative estimations between them and our own beliefs, desires, and so. At the end, the "idealized" character of rationality is given by the unavoidable choice of using ourselves as standards.

But Dennett's answer begs the question about the notion of rationality; he says nothing that allows a characterization of the attributed rationality. As a matter of fact, according to Dennett any attempt for defining rationality has to be done appealing to what we consider it is, to our "shared intuitions". The problem arises when one notice that most of our shared intuitions lead us to the conclusion

that rationality is defined in terms of logical consistency or deductive closure, precisely the terms Dennett is compelling us to avoid. However, those are some of the strongest intuitive notions of what to be rational would mean. When one is asked to be "as rational as possible" one is being required to be as consistent as one can be. Of course this is not the answer Dennett what us to give, but he has not offer us any alternative solution to such a request, and the requisite to use our intuitions lead us far from where Dennett wants us to be. It seems that Dennett is confusing the common sense and the theoretically necessary notions of rationality and makes use of each one of them when he need to, while treating them as incompatible when circumstances dictate. Additionally, when trying to escape from the temptation of "defining rationality in evolutionary terms," Dennett falls into the problem of treating as real something he defends to exist only as a pragmatic tool. As Stich has pointed out, Dennett's argument

> Slips almost unnoticeably from the claim that natural selection favors cognitive processes which yield true beliefs in the natural environment to the claim that natural selection favors *rational* belief-forming strategies. There are many circumstances in which inferential strategies which from a normative standpoint are patently invalid will nonetheless generally yield the right answer (1980, p. 53).

With his cryptic definition of rationality, Dennett cannot contend this sort of criticism, as it is impossible to resolve the puzzle Fodor and Lepore have unveiled.

5. Interpreting belief from the intentional stance.

Finally, there is another problem that threatens Dennett's approach to rationality and that his presentation does not resolve; the inevitable rationality of belief interpreters. In his attack on Quine's program for Natural Epistemology, J. Kim has introduced the problem of the notion of belief as normative on its own. Kim states that, given the non-normative character of natural epistemology, it is impossible to think about it as of being about beliefs. According to Kim:

> In order to study the sensory input-output relations for [a] given cognizer, we must find what "representations" he has formed as a result of the particular stimulations that have been applied to his sensory transducers. Setting aside the jargon, what we need to be able to do is to attribute beliefs, and other contentful intentional states, to the cognizer. But belief attribution ultimately requires a "radical interpretation" of the cognizer, of his speech and his intentional states; that is, we must construct an "interpretative theory" that simultaneously assigns meanings to his utterances and attributes to him beliefs and other propositional attitudes. […] Unless our cognizer is a "rational being", a being whose cognitive output is regulated and constrained by norms of rationality —typically, these norms holistically constrain his propositional attitudes in virtue of its contents— we cannot intelligibly interpret his "output" as consisting of beliefs. Conversely, if we

> are unable to interpret our subject's meanings and propositional attitudes in a way that satisfies a minimal standard of rationality, there is little reason to regard him as a "cognizer", a being that forms representations and construct theories. That means that there is a sense of "rational" in which the expression "rational beliefs" is redundant; every belief must be rational in certain minimal ways. […] Unless the output of our cognizer is subject to evaluation in accordance with norms of rationality, that output cannot be considered as consisting of beliefs (1988, p. 394).

Kim's account seems to be made to explain Dennett's methods and goals, and that is what justifies the long quote of his presentation. The central point of the presentation is the recognition of the necessity for the interpreter of behaviors, the cognizer, to be rational in order to be able to assign rationality to the agent whose behavior he is trying to make sense of. This is precisely Dennett's aim when stating the "idealized" character of his notion of rationality: no one can attribute rationality unless he himself is rational.

> Rationality in its broad and fundamental sense is not an optional property of beliefs, a virtue that some belief may enjoy and other lack; it is a precondition of the attribution and individuation of belief- that is, a property without which the concept of belief would be unintelligible and pointless (pp. 395-396).

What Kim is showing corresponds with what I want to point out as underlying Dennett's use of rationality attribution. The necessity of a notion of rationality possessed by the interpreter prior to undertaking his work introduces some doubts about the possibilities for finding in the observed subject something different from what we are using as an interpretative device. We do not just see what we say we see; what we see in the subject's behavior is precisely what we have introduced in the subject's mind in order to make sense of its behavior. When interpreting behaviors, it seems we are just seeing ourselves mirrored in the subject under scrutiny. Then what we find there is not surprisingly close to what we have, actually, it is just the image of our own mental content. If what we find and what we introduce are inescapably the same, then their equation is trivial and there is nothing to said to be found. On the other hand, if what is found is defined in terms of what is introduced, one needs to have a clear cut between them in order to avoid self-reference and infinite regress. Again, with the definition of rationality Dennett has provided us with such a difference is less than easy to be made.

Like in the case of the argument from natural selection, it has been shown that there are some open questions for Dennett's intentional attribution program in respect to its notion of ideal rationality and its attribution. Summed up, the questions presented here would not prove Dennett's approach to be fully misled or inconsistent; but this is not their purpose. Their main aim of the paper is to point out some aspects of the intentional attribution program that, if defensible, would require further developmentφ

BIBLIOGRAPHICAL SOURCES

Davidson, D. (1974). "Beliefs and the Basis of Meaning," in *Inquiries into Truth and Interpretation.* Oxford: Claredon Press.

Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.

Dawkins, R. (1986). *The blind Watchmaker*. Essex: Longman Scientific and Technical .

Dennett, D. (1971). "Intentional Systems," in *Brainstorms. Philosophical Essays on Mind and Psychology,* pp. 3-22. Montgomery: Bradford Book.

Dennett, D. (1978). *Brainstorms. Philosophical Essays on Mind and Psychology*. Montgomery: Bradford Book.

Dennett, D. (1979). "True believers," in *The Intentional Stance,* pp. 13-36. Cambridge, Mass.: MIT Press. 1987.

Dennett, D. (1987). "Error, Evolution and Intentionality," in *The Intentional Stance,* pp. 287-321. Cambridge MA: MIT Press.

Dennett, D. (1987). *The Intentional Stance*. Cambridge, Mass: MIT Press.

Dennett, D. (1996). *Kinds of Minds. Toward an Understanding of Consciousness*. New York: Basic Books. Harper Collins Publisher.

Dennett, D. (1998). *Brainchildren. Essays on Designing Minds*. Cambridge, MA: Bradford Book/ MIT Press.

Fodor, J. (1981). "Three Cheers for Propositional Attitudes," in *Representations*. Cambridge, Mass: Bradford Book/ MIT Press.

Fodor, J. and Lepore, E. (1993). "Is Intentional Ascription Intrinsically Normative?" in Dahlbom, *Dennett and his Critics. Demystifying Mind,* pp. 70-82. Cambridge, Mass: Blackwell.

Kim, J. (1988). "What is Naturalized Epistemology?" *Philosophical Perspectives 2 Epistemology*, 381-405.

Stich, S. (1981). "Dennett on intentional systems". *Philosophical Topics*, 12, 32-62.

Stich, S. (1984). "Could Man Be an Irrational Animal?" *Synthese*, 64, 115-135.